2020

Motor vehicle crashes and the urban built environment: A case study of a region in Des Moines

Dorcas Okailey Okaidjah
Iowa State University

Motor vehicle crashes and the urban built environment: A case study of a region in Des Moines

by

Dorcas Okaidjah

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF COMMUNITY AND REGIONAL PLANNING

MASTER OF SCIENCE CIVIL ENGINEERING

Majors: Community and Regional Planning; Civil Engineering (Transportation Engineering)

Program of Study Committee:

Monica Haddad, Co-Major Professor
Christopher Day, Co-Major Professor
Biswa Das

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this thesis. The Graduate College will ensure this thesis is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2020

DEDICATION

I dedicate this thesis to my positive and amazing parents (Florence and Benjamin Okaidjah), my loving husband (Richard Anarfi), and my beloved children that inspire me: Jayden Anarfi and Astrid Anarfi.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

AADT	Annual Average Daily Traffic
ESDA	Exploratory Spatial Data Analysis
GIS	Geographic Information System
GSV	Google Street View
HH	High High
HL	High Low
IDOT	Iowa Department of Transportation
LH	Low High
LISA	Local Indicators of Spatial Association
LL	Low Low
MV	Motor Vehicle
NB	Negative Binomial
RENB	Random Effect Negative Binomial
RMEV	Rate Per Million Entering Vehicles
TAZ	Traffic Analysis Zone

ACKNOWLEDGMENTS

My master's journey would not have been possible if it had not been for the Lord God Almighty, who ordered my steps and granted me peace, wisdom, courage, and strength. I would also like to express my utmost gratitude to Prof. Monica Haddad, who has been instrumental in my studies and offered me guidance throughout this journey. I would like to immensely thank Prof. Christopher Day, who accepted to be my major professor in civil engineering. Without whose help, I could not have added on a civil engineering master's degree. I want to thank Prof. Biswa Das for being on my committee and offering me advice and support. I would also like to thank Prof. Carlton Basmajian, whose class (a mix of history and future of planning, stories, thoughtful discussions, humor, and food; exactly what I pictured graduate teaching to be like) I enjoyed and deepened my interest in urban planning. Prof. Carlton Basmajian also assisted me all through the process and paperwork in adding on a civil engineering master's degree and tolerated my many questions, for which I am immensely grateful. I would also like to thank Prof. Francis Owusu, whose advice I sought when I needed to make an educated decision and whose Christmas and random parties I enjoyed. I am tremendously grateful to my best friend, Richard Anarfi, who has encouraged and supported me through my graduate studies. I would also like to thank all the wonderful colleagues who became friends along this journey, the department faculty, and staff for adding color to my time at Iowa State University.

ABSTRACT

Existing scholarly work demonstrates that the built environment can affect the frequency of motor vehicle (MV) crashes. The objective of this study is to explore the relationship between urban MV traffic crashes at road intersections and the built environment in the city of Des Moines. The study area includes low-income and wealthy neighborhoods to understand the built environment in these different contexts. Exploratory Spatial Data Analysis (ESDA) is used to identify MV crash hotspots at intersections. Google Street View (GSV) is used as a tool to survey the built environment variables such as commercial/institutional land uses, schools, parks, signage, number of lanes, on-street parking, bus stops, etc. of the hotspot intersections. Multiple linear regression and a Tobit model is then employed to establish a relationship between MV traffic crash hotspots at intersections and the built environment.

The study considers the statistical significance of the MV crash locations; hence it employs an exploratory spatial data analysis in analyzing MV traffic crashes. This exploratory case study was conducted using 7-year data of vehicle crashes from 2013 to 2019 obtained from the Iowa Department of Transportation (IDOT). The study results indicate that commercial/institutional land uses, bus stops, and signalized intersections are significant built environment variables that impact the occurrence of MV traffic crashes. Additionally, the results also show that MV traffic crash hotspot intersections were in areas with a high household poverty percentage. These results can inform policymakers to develop strategies that focus on suitable MV traffic safety, such as traffic calming measures in hotspot locations. Design ideas to improve the built environment and a policy framework for bus stop locations can be developed, thus preventing and reducing MV traffic crashes.

CHAPTER 1. INTRODUCTION

1.1 Study Background

Traffic crashes pose a concern locally and nationally due to the loss of human lives and property damage and account for one of the leading causes of death in the United States (Hoyert et al., 2005). To illustrate, in 2010, motor vehicle crashes resulted in 33 thousand people killed, 3.9 million injured, and 24 million vehicles damaged in the United States (Blincoe et al., 2015). In 2018, an estimated number of 40 thousand fatalities occurred owing to motor vehicle crashes and 4.5 million injured in the United States (National Safety Council, 2020).

The economic costs of these 2010 crashes totaled $242 billion, encompassing losses in labor productivity, medical, legal, and court costs, and expenses related to emergency service, insurance administration, congestion, property damage, and workplace losses (Blincoe et al., 2015). Additionally, in 2018 medically consulted injuries in motor vehicle incidents totaled 4.5 million, and economic costs were estimated at $445.6 billion. The financial cost of crashes can be mitigated with safety strategies in transportation planning (National Safety Council, 2020).

The state of Iowa is no exception to the problem relating to traffic crashes and loss of human lives. In 2018, the total motor vehicle crashes were 56 thousand, with 291 fatal crashes,15 thousand injury crashes, 18 thousand injuries, and 319 fatalities. The capital city of Iowa, Des Moines, has experienced an increase in traffic crashes by 23.8% from 2013-2017 and a decrease of 2% between 2017 and 2018 (Iowa Department of Transportation, 2020).

The built environment has been defined as a broad concept that includes land use patterns, transportation systems, and design features that allow for travel and physical activity (Transportation Research Board, 2005). Additionally, the built environment comprises urban design elements, land use, and the transportation system that factors human activity patterns within the physical environment (Handy et al., 2002). For this study, I defined the built environment as physical features visible in the urban environment: land use, parking, sidewalks, bus stops, street design elements, signage, etc.

The built environment contributes to traffic-related deaths and injuries, influencing the frequency of crashes and its severity hence the need to study the built environment (Ewing & Dumbaugh, 2009). More so, built environment attributes appear to play a profound role in the occurrence of traffic crashes (Dumbaugh & Li, 2010). In a study of the relationship between built environment attributes and severity of the crash injury, Sze and Wong (2007) found that the number of lanes and the occurrence of crashes at crosswalks or signalized intersections led to a higher risk of mortality and injury severity. Kaygisiz et al. (2017) studied the relationship between the built environment and vehicle-vehicle crash frequencies. The study's findings indicate that mixed land uses contributed to such crashes; hence safety improvements should concentrate on places with varied land uses. The studies suggest that the built environment is an essential component of traffic safety.

Understanding the spatial and temporal crash patterns of a city provides traffic safety engineers and planners information to detect the transportation network sections having a higher number of crashes. Sections with a higher number of crashes are referred to as hotspots (Elvik, 2008). Therefore, identifying vehicle crash hotspots in urban areas is essential to gaining a comprehensive understanding of vehicle crashes and the built environment.

1.2 Research Objectives and Questions

The study's main objective is to understand the relationship between MV traffic crashes at intersections and the built environment in the City of Des Moines. Specifically, I identify intersections with high and low occurrences of motor vehicle traffic crashes and describe the built environment variables that surround these intersections. I pose the following spatial questions: where are the locations of high and low clusters of urban MV traffic crashes intersections during 2013-2019 in the City of Des Moines? What visual elements of the built environment characterize the high and low clusters of MV traffic crash intersections? What is the relationship between high and low clusters of MV traffic crash intersections and the level of poverty in the neighborhoods they are located?

1.3 Significance of the Research

The research is significant because many studies focus on the relationship between built environment and pedestrian crash counts as well as pedestrian safety and injuries (Clifton et al., 2009, Zahabi et al., 2011 & Stoker et al. 2015). The injury severity in pedestrian vehicle crashes has been examined considering personal and environmental variables such as land use, transit access, population density, weather, pedestrian location, etc. (Clifton et al., 2009). Also, accidents at intersections have been shown to influence pedestrian injury severity (Zahabi et al., 2011).

However, very few studies investigate the impact of the built environment on MV traffic crashes. Additionally, existing studies on MV traffic crashes tend to focus on the regional level and not on the intersection level; hence, this study's uniqueness in focusing specifically on local, city-based intersections where there is a high occurrence of MV traffic crashes. This study investigates the relationship between the built environment and MV traffic crashes at intersections due to the lack, as mentioned earlier, of literature. Also, it considers the influence of the socioeconomic variable poverty in the location of MV traffic crashes, of which there is very little literature.

This research study reduces the possibility of accidents due to inclement weather conditions such as the winter season to emphasize the built environment variables. Therefore, the research study is significant to understanding what built environment variables influence MV traffic crashes at intersections, enabling transportation planners and engineers to know what kind of measures to put in place to mitigate the occurrence of such crashes. This research would be useful in land use planning and enhancing collaboration between transportation planners and transportation engineers.

1.4 Thesis Organization

The thesis is organized as follows. Chapter one is the introduction, which entails the study background, research objectives and questions, and the study's significance. Chapter two encompasses previous studies on traffic crashes and low-income neighborhoods, traffic crashes and the built environment, empirical studies examining traffic crashes, and empirical studies on the use of Google Street View (GSV). Chapter three entails the study area, describes the data used in the study and provides information on the accident trends in the City of Des Moines. Chapter four includes information on the methodology, and Chapter five details the results from analyzing the relationship between MV traffic crashes and the built environment. Chapter six discusses the research findings, recommendations, limitations, and future research.

CHAPTER 2. REVIEW OF LITERATURE

This chapter focuses on four main topics: (1) traffic crashes and low-income neighborhoods; (2) traffic crashes and the built environment; (3) empirical studies on traffic crashes, and (4) empirical studies on Google Street View (GSV) for virtual audit. The first topic includes preceding studies that investigate the relationship between traffic crash locations and low-income areas. The second topic gives background information on previous studies that examine the characteristics of the built environment variables that impact traffic crashes. The topic of traffic crashes focuses on methods and models used in previous research to analyze traffic crashes, and the final topic concentrates on studies that show that google street view can replace in-situ fieldwork

2.1 Traffic Crashes and Low-Income Neighborhoods

Cottrill & Thakuriah (2010) undertook a research study evaluating pedestrian crashes in areas with high-low income or minority populations in the Chicago metropolitan area using a Poisson model. Findings indicated that low-income neighborhoods, typically with minority populations, have high crashes, especially pedestrian crashes. Variables of exposure to crashes included walkability of the area, transit accessibility and availability, crime rates, and general population demographics such as income and presence of children. As a result, safety improvements should be implemented within high-low income or minority population areas. Dezman et al. (2016) conducted a study to investigate traffic crashes' locations and their characteristics in the city of Baltimore. Crash and socioeconomic data were used for this study. Exploratory spatial data analysis and spatial autocorrelation were employed to examine the geographical distribution of traffic crashes and hotpots. The influence of socioeconomic indicators on hotspots was evaluated using spatial regression. Findings indicated a high occurrence of road traffic crashes in high density locations of Baltimore city and an upsurge of crashes from March to June. These crashes were also characterized by inattentive driving. However, socioeconomic variables were not associated with crash occurrence and hotspots.

Morency et al. (2012) studied social inequalities in motor vehicle, pedestrian, and cyclist injuries across wealthy and poor urban neighborhoods considering the influence of traffic volume and road geometry. The study findings showed that significant numbers of pedestrians, cyclists, and motor vehicle occupants' injuries existed at intersections in the most impoverished areas compared to intersections in the wealthiest areas.

2.2 Traffic Crashes and the Built Environment

Ouyang & Bejleri (2014) analyzed traffic crashes at the community level in Miami-Dade County, Florida. The study employed a geographic information system to process crash and built environment data and a negative binomial (NB) model regression to assess the influence of the built environment on traffic crashes. Spatial and non-spatial data were combined and analyzed in geographic information system (GIS) with the census block group as the spatial unit, which is classified as being on a community scale. The built environment variables were typified based on aspects of density, diversity, design, destination accessibility, and distance to transit. Findings of the study indicated that mixed land use and the number of bus stops was positively associated with all crash types, that is, property damage only, injury and fatal crashes, and the number of intersections and density has little influence on crash types.

According to Wolf & Bratton (2006), community character and environmental systems that characterize the roadside have been ignored in the transportation planning process; hence a study was conducted to investigate national traffic collision data to address issues about urban trees and traffic safety encompassing crash occurrences and severity. The study findings recognize that tree crashes have profound implications, and it must not be assumed that trees are the cause of crashes, but other intricate circumstances, that is, driver or road environment-related factors, can contribute to crash occurrence. The study concludes that tree crashes in urban settings are not explicitly understood hence the need to investigate further (Wolf & Bratton, 2006).

Gladhill & Monsere (2012) argued that the urban form influenced traffic safety and employed an NB regression to model crashes to understand the extent of the influence. The variables defining the urban form were demographics, trips, street layout, exposure, connectivity, and transit accessibility. The study area was Portland, Oregon, and the study also modeled crashes by transit mode, crash type, and injury severity. Study findings showed that business density, population, and transit stops were significant variables in most crash models. Intersection density and street connectivity were not substantial at varying crash severity levels, which contradicted earlier studies.

Huang et al. (2018) investigated the relationship between vehicle crashes and the built environment using a geographically weighted regression (GWR) model. The study area was in the Detroit region of Michigan. The study's findings indicated that built environment variables, such as commercial use percentage, local road mileage percentage, and intersection density, have a relatively constant relationship with crashes. A higher percentage of commercial uses is linked to a higher crash rate, and four-way intersection density in a block group is typically associated with a higher crash rate. The findings also conclude that the relationships between the built environment and crashes are spatially non-stationary: both the strength and direction of their relationships differ over space.

Kim et al. (2006) explored the relationships between land use, population, employment by sector, economic output, and motor vehicle accidents employing various linear regression models. The NB model is used to examine the relative effects of land use, population, employment by sector, and economic output on the number of pedestrians, bicycle, vehicle-to-vehicle, and total accidents. A multivariate model is used to analyze the relationships between zonal characteristics and accidents. Study findings show that vehicle-to-vehicle crashes were associated with high employment and commercial activities.

Pulugurtha et al. (2013) researched to develop crash estimation models at the traffic analysis zone level based on land use characteristics. Findings indicated that land uses that generated high activities were positively associated with crashes in a traffic analysis zone. Land uses classified as high activity generators in this study included "mixed-use development area (with residential and compatible non-residential uses less than 10 acres to serve the residents of the planned community), urban residential area (nearer to the employment core), single-family residential area (with 3–8 dwelling units per acre), multi-family residential area (with 12–43 dwelling units per acre), business area (with large trade area, warehousing, wholesaling, etc.) and office district area."

In a study by Dumbaugh & Li (2010), a high percentage of urban crashes in urban environments were not random but appeared to be influenced by the built environment's characteristics. The study concluded a positive association between crashes and attributes of the built environment. Major crash risk factors identified in the study were commercial strip uses, big box stores, miles of arterial roadways, and numbers of four-leg intersections.

Based on the literature on traffic crashes and the built environment, a research hypothesis can be formed. The presence of commercial activities and high density centers could contribute to the occurrence of traffic crashes.

2.3 Empirical Studies-Traffic Crashes

Traffic crash hotspots identification has been studied and implemented in previous literature using various methods (Dai & Jaworski 2016; Truong & Somenahalli 2011; Xie & Yan, 2008). Kernel density estimation is a popular method used to determine traffic crashes (Anderson, 2009; Flahaut et al., 2003). The two kinds of kernel density estimation used in literature are the Planar Kernel Density Estimation (PKDE) and Network Kernel Density Estimation (NKDE).

The PKDE method results in a smooth density surface of spatial point events over a 2-D geographic space, and it is less frequently used in traffic crash studies. The PKDE method is unpopular because of its likelihood of overestimating the density values and covering space beyond the transportation network space (Xie & Yan, 2008; Benedek et al., 2016). The NKDE Method is more frequently used than the standard PKDE since the density of events is calculated strictly along with the transportation network (Xie & Yan, 2008). The challenge of using planar or network kernel density estimation is the inability to test for statistical significance (Xie & Yan, 2008).

However, Local Indicators of spatial autocorrelation (LISA) is a method that indicates the presence of significant spatial clusters or outliers for each location (Anselin, 1995). Truong & Somenahalli (2011) performed research using Moran's I to examine the spatial patterns of pedestrian-vehicle crashes and the Getis-Ord Gi statistic to identify pedestrian-vehicle crash hotspots. A cluster of high index values was represented by a high value of Getis-Ord Gi statistic and vice-versa. In summary, many methods have been used to analyze traffic studies to determine traffic crash hotspots and to understand the spatial patterns of such crashes.

Several statistical regression methods have been employed in traffic crashes analysis. The NB model was used by Gladhill & Monsere (2012) to explore traffic safety and the urban form and by Ouyang and Bejleri (2014) to understand the effects of the built environment on traffic crashes. Kim, Brunner & Yamashita, 2006 employed both linear regression and NB regression models to explore the relationship between land use, population, employment by sector, economic output, and motor vehicle accidents.

Chin & Quddus (2003) employed a random effect negative binomial model (RENB) to investigate the relationship between crash occurrence and the geometric, traffic, and control attributes of signalized intersections. The RENB model's suitability over the Poisson or negative binomial NB model is because of unobserved heterogeneity and serial correlation in the crash data.

Anastasopoulos, Tarko & Mannering (2008) used the Tobit model to examine vehicle accident rates left-censored at zero using vehicle accident data from Indiana interstates. The model showed pavement conditions, roadway geometrics, and traffic characteristics impact vehicle accident rates significantly. Park & Lee (2017) developed a random parameter Tobit model that considers heterogeneity in accident data and analyzes nine years of severe injury accident data in Washington State, USA. The results showed that the random parameter Tobit regression model provided a better understanding of factors impacting severe injury accident rates than the fixed parameter Tobit model. The random parameter considers varying parameter observations, while the fixed parameter considers constant parameter observations.

2.4 Empirical Studies-Google Street View

Rundle et al. (2011) performed research assessing the feasibility of using Google Street View (GSV) to audit neighborhood environments. The study utilized 2007 neighborhood audit data to examine neighborhood measurements coded in 2008. 37 block faces in New York City's high-walkability neighborhoods were among the sample. Field audit and GSV data were gathered for 143 items related to seven neighborhood environment characteristics: aesthetics, physical disorder, pedestrian safety, motorized traffic and parking, infrastructure for active travel, sidewalk amenities, and social and commercial activity. Percentage agreement for categorical measures was employed to measure the relationship between field audit and GSV data. The study revealed that GSV could be used to audit neighborhood environments.

Kelly et al. (2013) undertook research to access the inter-rater reliability audits using GSV imagery. In 2011, street segments from St. Louis and Indianapolis were geographically stratified to represent neighborhoods with different land use and socioeconomic characteristics. Inter-rater reliability was assessed using observed agreement and the prevalence-adjusted bias adjusted kappa statistic (PABAK). The research concluded that the use of GSV to audit the built environment is a dependable method for analyzing the built environment's characteristics.

Badland et al. (2010) performed research to analyze if virtual streetscape audits can reliably replace physical streetscape audits. Built environment characteristics linked with walking and cycling were examined using the New Zealand Systematic Pedestrian and Cycling Environment Scan (NZ-SPACES) in 48 street segments. These street segments were from four neighborhoods in Auckland, New Zealand. Audits were performed physically (on-site) and remotely (using GSV) in January and February 2010. The findings demonstrated that GSV was, for the most part, an efficient and effective tool to measure the streetscape context at the neighborhood level. Mooney et al. (2016) performed research using GSV to assess the environmental contribution to pedestrian injury. GSV's imagery from 2007 to 2011 was used to examine 9 characteristics of 532 intersections within New York City. Consistent with in-person study observations, the information-technology approach found traffic islands, visual advertising, bus stops, and crosswalk infrastructures to be linked with elevated counts of pedestrian injury in New York City.

Clarke et al. (2010) performed research to capture community characteristics on the physical and mental health of residents; hence a virtual neighborhood audit was conducted by the city of Chicago using GSV. Data from GSV was compared to in-person data that was also gathered from an identical audit. Findings indicated that reliable data of community characteristics such as general land use, recreational facilities, and the local food environment could be obtained from a virtual audit instrument. The study, however, stated to organize caution when trying to gather finely detailed observations.

Odgers et al. (2012) conducted a study to find out whether neighborhood conditions influence behavior and health. The method employed for the study was a virtual systematic social observation (SSO), to assess whether GSV could reliably capture the neighborhood conditions. Negative neighborhood features, including disorder and decay and dangerousness, coincided with reports from local resident reports and socioeconomic status data from the census. Positive neighborhood features, street safety, and the percentage of green space were associated with higher prosocial behavior and healthy weight status among children. The study indicated that using GSV is a reliable, cost-effective measure of capturing positive and negative neighborhood features. In summary, google street view has been shown to be a dependable method to audit built environment data, and GSV has been shown to be in concordance with field audit (Mooney et al., 2016; Badland et al., 2010; Kelly et al., 2013; Rundle et al., 2011).

2.5 Unit of Analysis

Block groups and census tracts are widely used as the unit of analysis in crash analysis and are easily integrated with socioeconomic and demographic census data. The smallest geographic units used by the U.S. Census Bureau are census blocks. A group of census blocks is known as block groups and, several census blocks are integrated into census tracts. Ouyang & Bejleri (2014) evaluated the built environment and traffic crashes employing the census block group as the spatial unit of analysis and aggregated housing and population data at the census block group. Dezman et al. (2016) investigated the causes of motor vehicle crashes using the census tracts to conduct data analysis and to merge socioeconomic data.

Traffic analysis zones (TAZ) have been employed as a unit of analysis (Soltani & Askari, 2017 & Pulugurtha et al.,2013) for crash studies. TAZ comprises one or more census blocks, block groups, or census tracts that integrates land use, socioeconomic and sociodemographic characteristics to estimate trip generation and attraction (Meyer & Miller, 2001). The use of TAZ is preferred because of the availability of crash data at the macro-level (Lovegrove & Sayed, 2006).

Wang & Kockelman (2013) studied the relationship between 3-year pedestrian crash counts across census tracts in Austin, Texas, and diverse land use, network, and demographic characteristics. The demographic characteristics included accessibility, Job and resident densities, lane-mile and sidewalk densities, and land use balance. The unit of analysis was Thiessen polygons constructed around Austin's census tracts centroid. The 3-year crash counts were aggregated over Thiessen polygons employing the ArcGIS's spatial join function. Wang & Kockelman (2013), in their study, recommends the aggregation of area level crash count data over Thiessen polygons to properly assign high crash locations to a polygon instead of randomly assigning to or split across the neighboring tract.

Kim et al. (2006) employed grid cells of the same size and shape as the analysis unit to explore the relationships between land use, population, employment, economic activity, and motor vehicle accidents. The grid cells were used to avoid the problem of different geographies associated with using census block groups, zip-code level employment data, statewide land use data, and countywide zoning data.

CHAPTER 3. STUDY AREA AND DATA DESCRIPTION

3.1 Study Area: City of Des Moines

Des Moines is the capital of Iowa and the most populous city in the state, with the county seat being Polk County. The Des Moines population was at 215,932 in 2018 and has experienced a population growth rate of 6.3% from 2010 to 2017. The median household income is $52,25, and the median age is 33.9 currently. (U.S Census Bureau, 2020). The study area chosen within Des Moines captures both low-income and wealthy neighborhoods to understand the differences in the built environment of the hotspot's intersections of MV traffic crashes. Poverty in the United States refers to an individual earning less than $34 per day or a family of four earning less than $69 per day. It is calculated from the poverty threshold defined by the United States Census Bureau (Federal Safety Net, 2020). The number of neighborhoods in the study area is 21, as shown in figure 3.1. Figure 3.2 is a choropleth map of the percentage of household poverty.

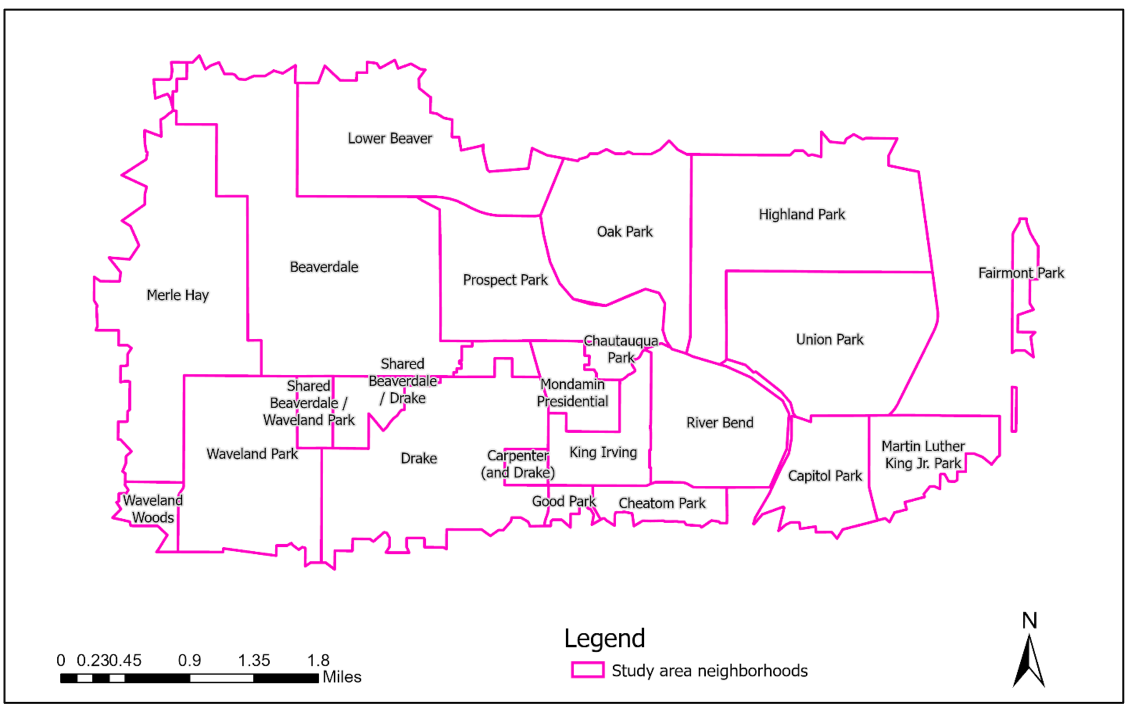

Figure 3.1: Study area neighborhoods

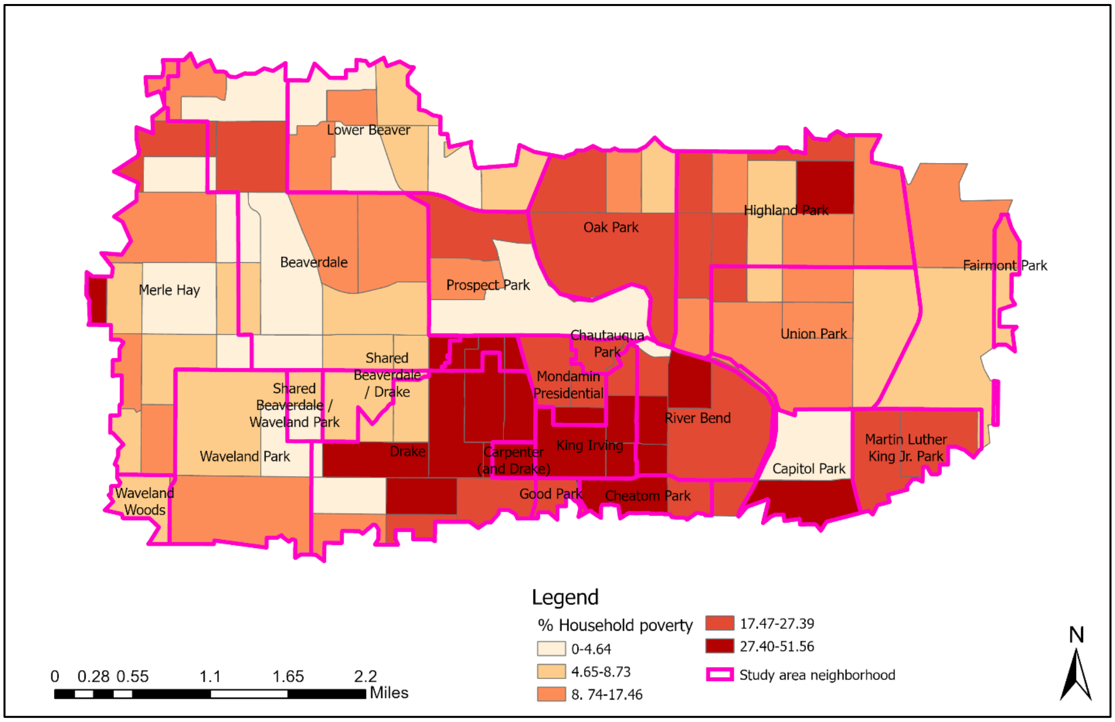

Figure 3.2: Choropleth map of household poverty in the study area

3.2 Accident Trends in Des Moines

Motor vehicle crashes in Des Moines have been increasing from 2011 to 2016 by 27.7%. A decrease of 2.7% in 2017 and 1.9% in 2018 was experienced, which might be due to safety efforts by the IDOT. However, the crashes increased again in 2019 by 4.4% (Iowa Department of Transportation, 2020). Figure 3.3 displays the trends in motor vehicle crashes in Des Moines and the USA from 2009 to 2018. The USA has a higher increasing trend in motor vehicle crashes as compared to Des Moines. Des Moines has seen upward and downward trends in serious and minor injuries and fatalities over the last decade, as shown in figure 3.4.

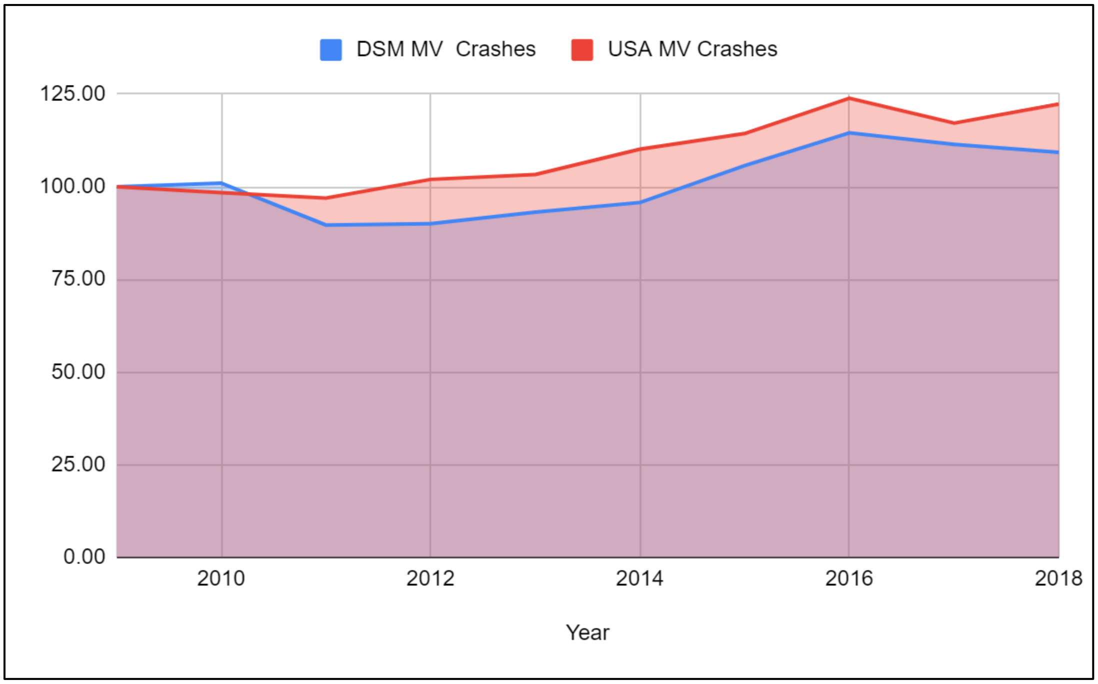

Figure 3.3: Des Moines and USA MV crashes (2009-2018)

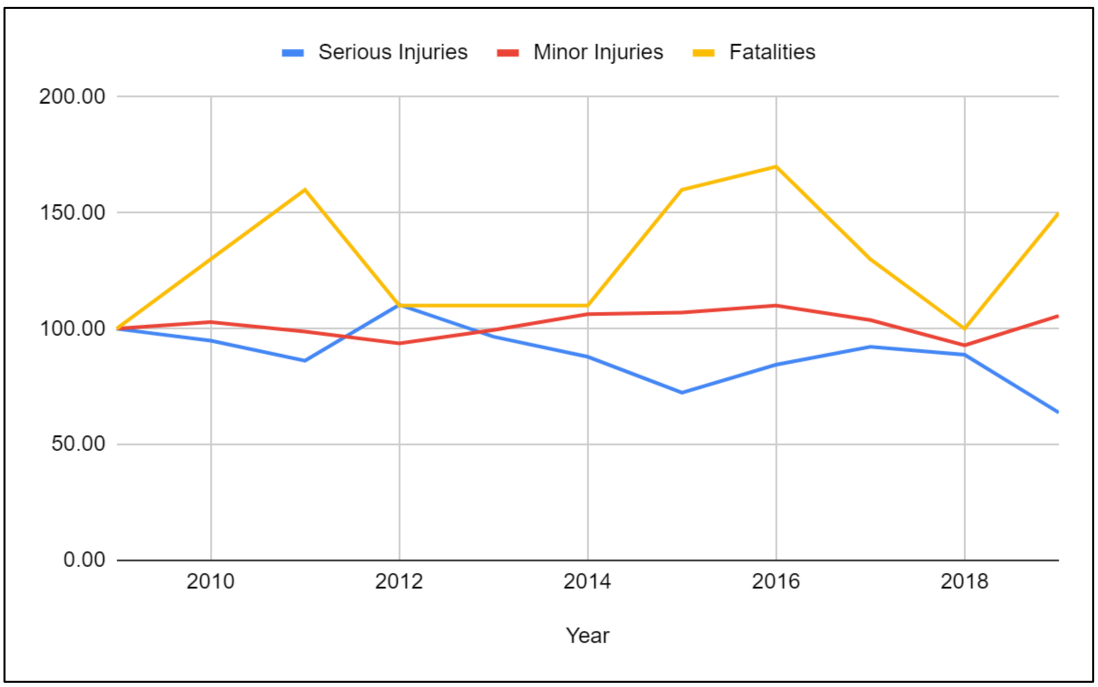

Figure 3.4: Des Moines crash severity (2009-2019)

3.3 Transportation Data

This study's transportation data includes the motor vehicle crash data and the traffic volume data that is the Annual Average Daily Traffic (AADT). The motor vehicle crash data were obtained from the IDOT website. It includes the case number, date of crash occurrence, crash location (the x coordinate and the y coordinate), and the crash severity.

The crash data used for the study was between the years 2013 and 2019, summing up to 7 years. This study filters out months with a high occurrence of snow, i.e., winter months. The months filtered out were January, February, March, and December. The crash data for the study area after January, February, March, and December were filtered out is displayed in figure 3.5, and figure 3.6 displays the crash data for the entire city of Des Moines. Hence, for this study, eight months of motor vehicle crash data were used. The months filtered out were to reduce accident happenings because of snow, ice, or bad weather.

The Annual Average Daily Traffic (AADT) represents the total volume of vehicles that passed a specific point in a year divided by 365. It is essential data used in transportation planning and analysis. The AADT was obtained from the IDOT website and was used to calculate MV intersection crash rates.

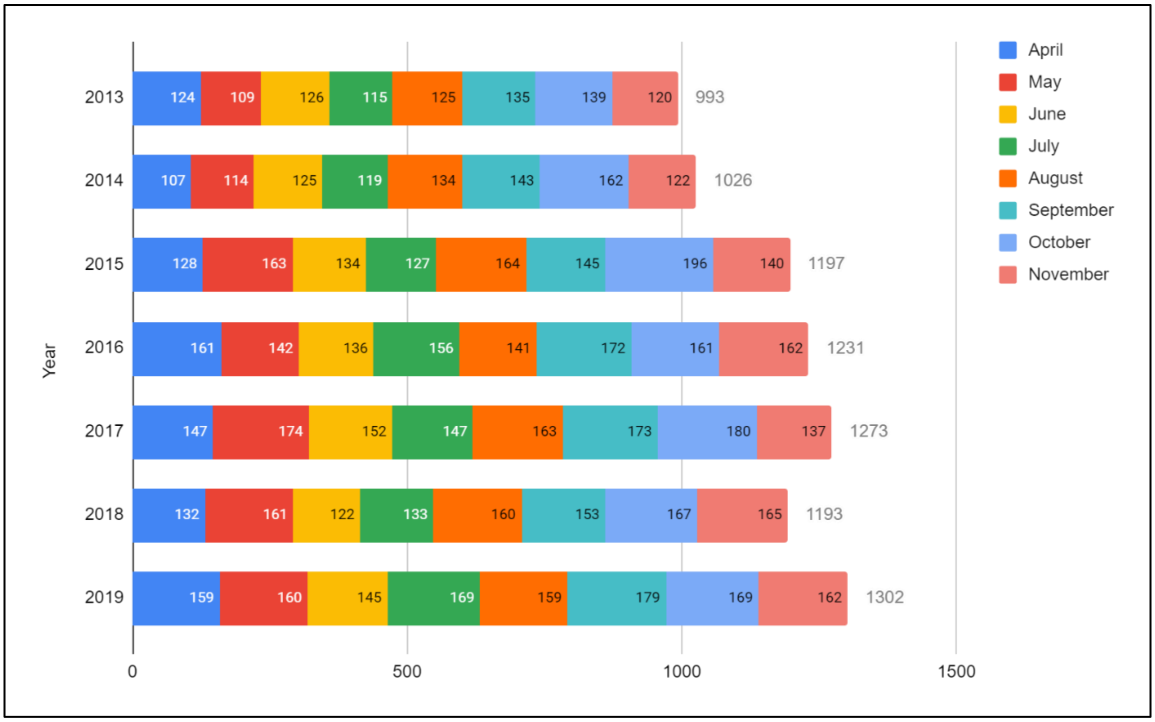

Figure 3.5: Study area MV traffic crashes (2013-2019)

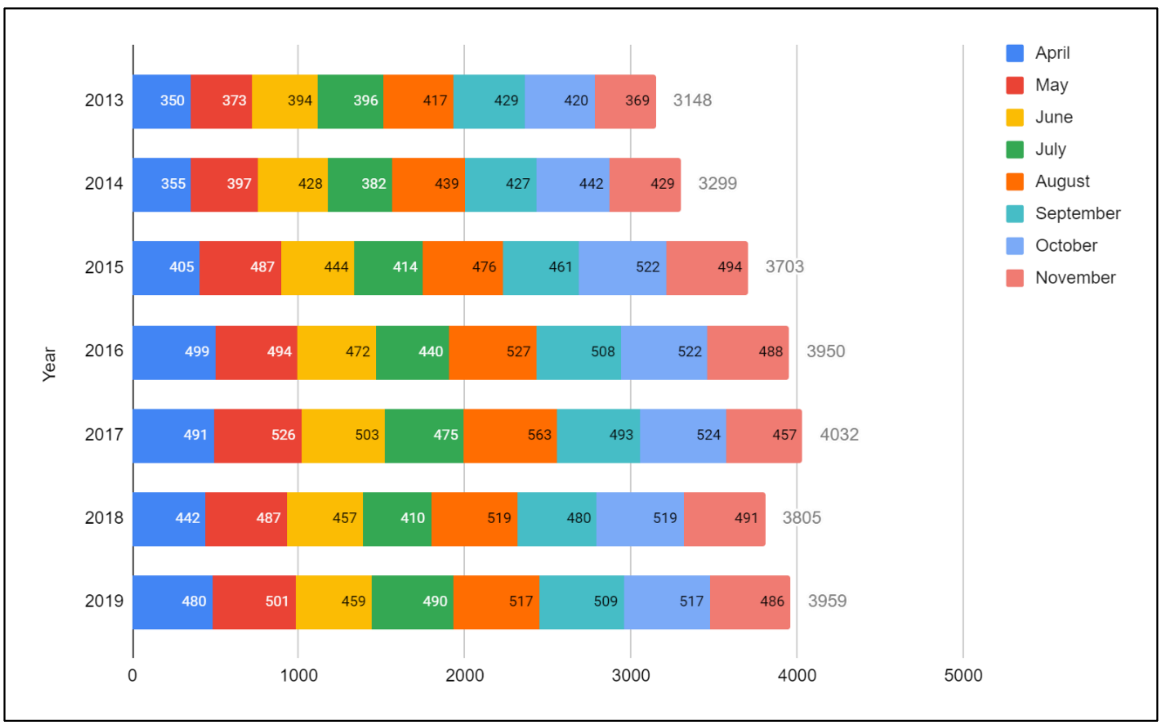

Figure 3.6: Des Moines MV traffic crashes (2013-2019)

3.4 Spatial Data

The following spatial data were obtained and used for this study: City of Des Moines boundary shapefile, Des Moines neighborhoods shapefile, highway shapefile, roads shapefile, census block groups, and household poverty data. The spatial data was obtained from the City of Des Moines and the Iowa Department of Transportation website. The census block groups, and household poverty data were obtained from the United States Census Bureau.

CHAPTER 4. METHODOLOGY

As stated in the introduction, the three research questions of this study are: where are the locations of high and low clusters of urban MV traffic crash intersections during 2013-2019 in the City of Des Moines? What visual elements of the built environment characterize the high and low clusters of MV traffic crash intersections? What is the relationship between high and low clusters of MV traffic crash intersections and the level of poverty in the neighborhoods they are located? This chapter elaborates on the methods employed in this research, providing information on the study design, methodological steps, and analytical methods used in the study.

4.1 Study Design

The study area includes low-income and high-income neighborhoods to investigate if MV traffic crashes are associated with the level of poverty. Exploratory Spatial Data Analysis was employed to detect locations with high and low clusters of traffic crashes. Google Street View questionnaire was used to gather data on the built environment's characteristics at high and low clusters of MV traffic crash intersections. These built environment variables were aggregated from previous studies that examined the built environment's influence on traffic crashes. The relationship between built environment variables and the high and low clusters of MV intersection traffic crashes was explored using multiple linear regression and the Tobit model. The methodological framework of the study is shown in figure 4.1. Results from the methodological steps will inform policymakers about land use decisions based on significant built environment variables identified.

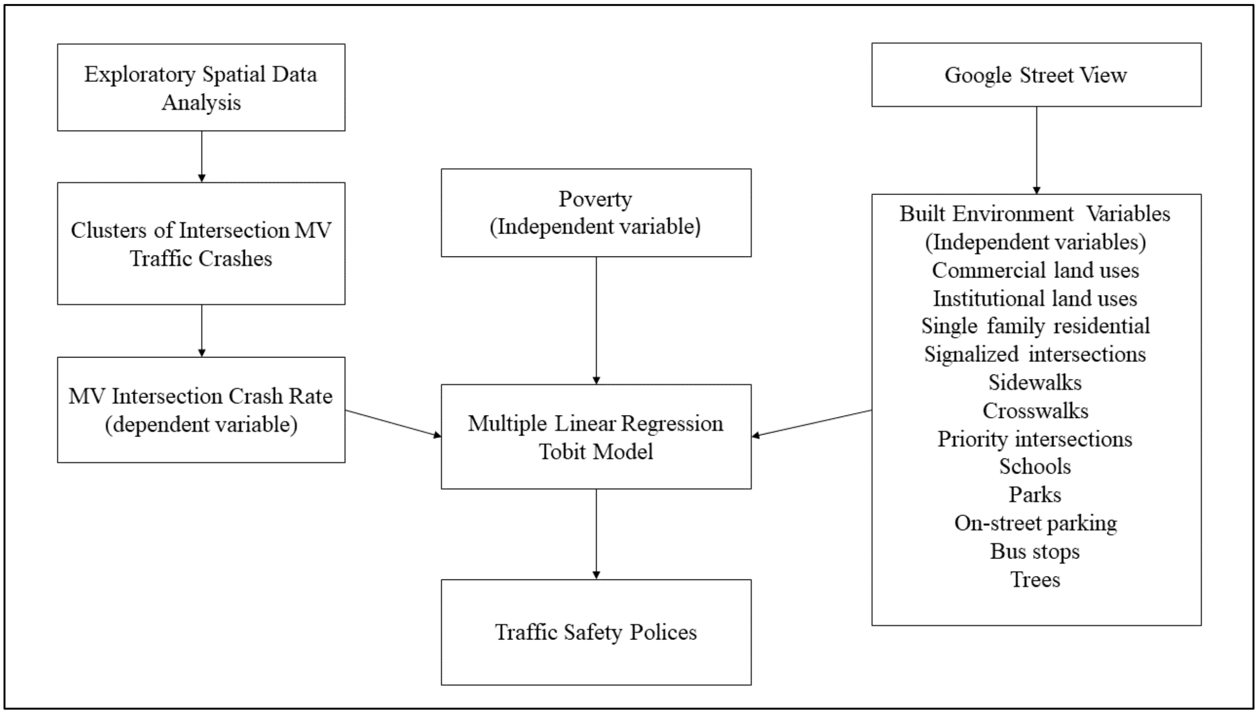

Figure 4.1: Methodological framework

The methodological process employs quantitative methods that involve manipulating pre-existing statistical data using computational techniques. The study uses quantitative data for analysis, which is obtained based on precise, objective measurements using structured and validated data-collection instruments. The data used includes traffic crash data and GSV questionnaire data. The GSV questionnaire was based on the observation of built environment variables elicited a yes or no response hence falls under quantitative data.

Quantitative analytical methods employed in this study entails spatial and statistical analysis. The spatial analysis involves mapping counts of traffic crash data and identifying clusters of MV traffic crashes, while statistical analysis includes the modeling of MV intersection crash rate and built environment variables.

4.2 Data Cleaning

Data cleaning and organization was necessary to begin the data processing for exploratory spatial data analysis. The study focuses on roads in the neighborhoods where the built environment variables can be observed; however, the interstate highway (I- 235) is at the border of the study area; hence it was necessary to clean it. Additionally, I-235 accommodates the highest traffic volumes in Iowa, with a daily average between 75,000 to 125,000. Trips made on the interstate highways are usually long-distance travel, with a significant proportion contributing to external trips.

The highway network data of Iowa was obtained and clipped to the city of Des Moines, as shown in figure 4.2, which also indicates I-235 and the study area within the entire City of Des Moines. I -235 was then selected from the highway network data and cleaned from the study area. Interstate highways such as I-235 tend to have forgiving design features such as wide lanes, wide shoulders, and clear roadside zones that influence a reduction in the crash incidence (Dumbaugh & Li 2010). There is little to no built environment variables to observe along the interstate highway due to roadside clear zones' existence, hence its removal from the study. The study area roads have very little or no wide lanes, wide shoulders, or clear roadside zones

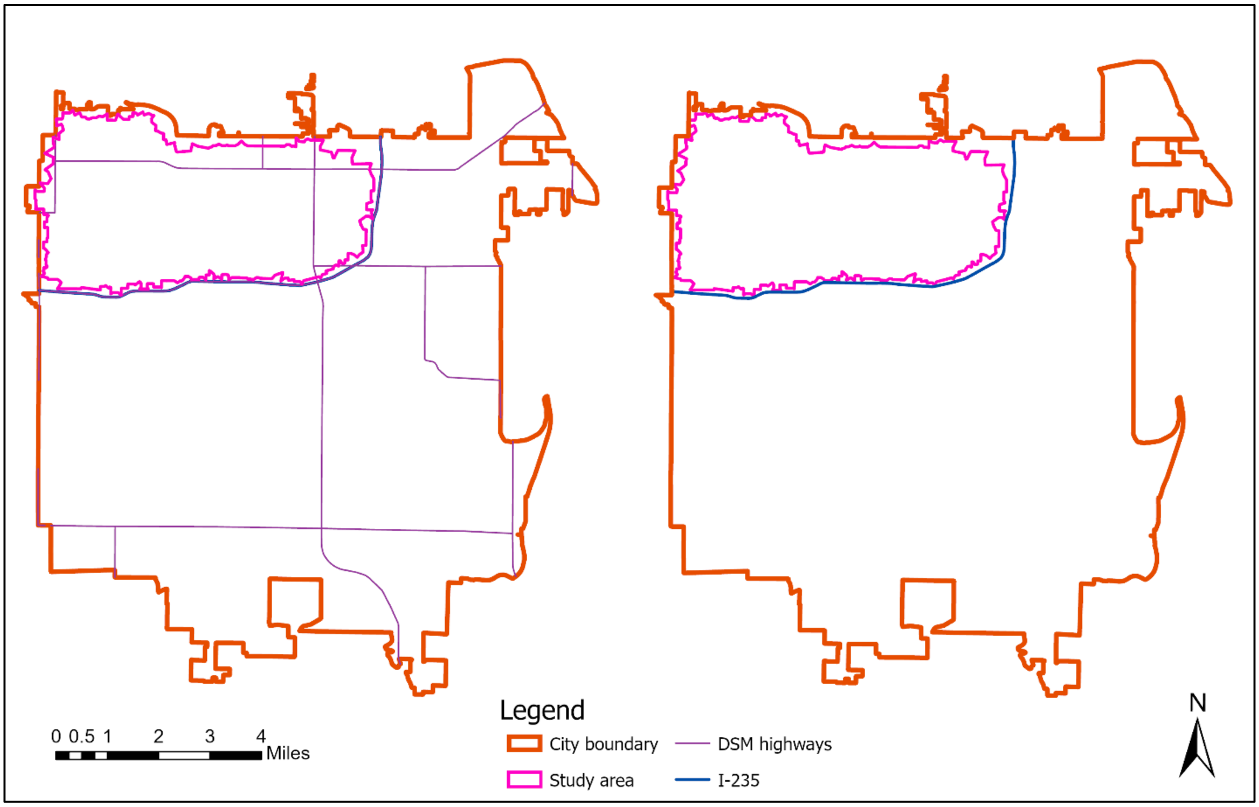

Figure 4.2: Study area data cleaning

A Python code was developed to clean intersections points that were too close together, within 66ft or less. This was necessary because concentric circles would pose a challenge when analyzing observations in the Thiessen polygons. The data for the python code processing included all intersection points in the study area with the x and y coordinates. The intersection points were filtered based on Euclidean distance. Figure 5.2 and 5.3 illustrates the intersection points before and after the python code was implemented. The python code is shown in appendix B.

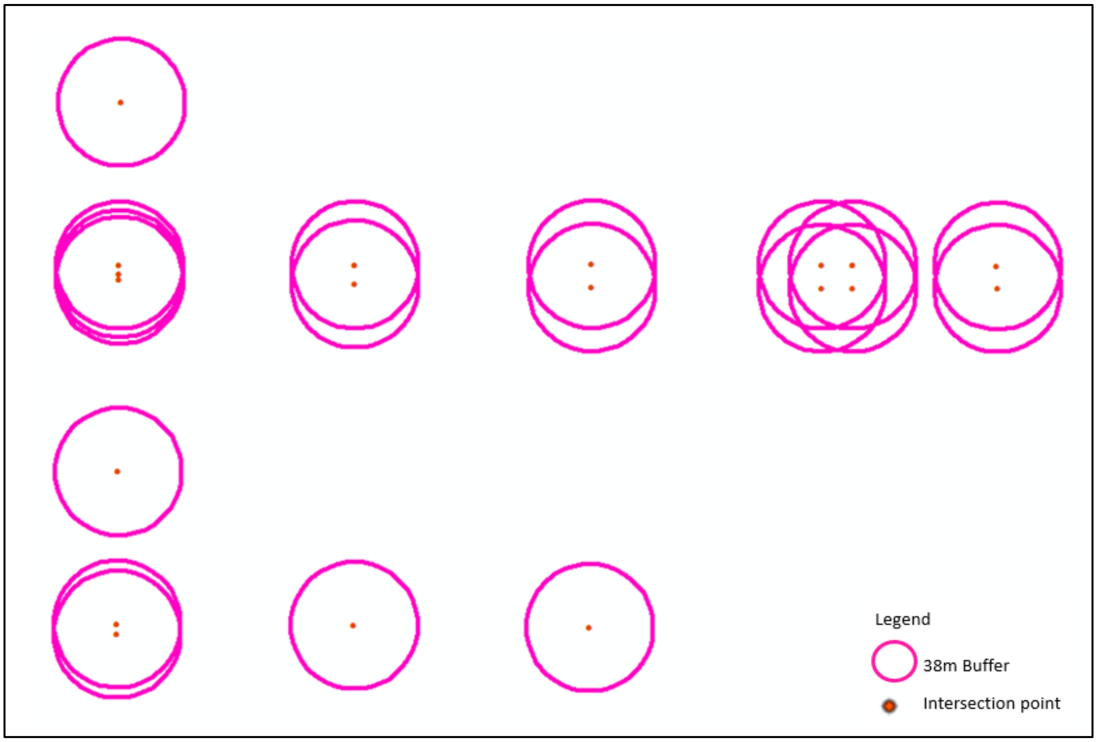

Figure 4.3: Buffered intersection points before data cleaning

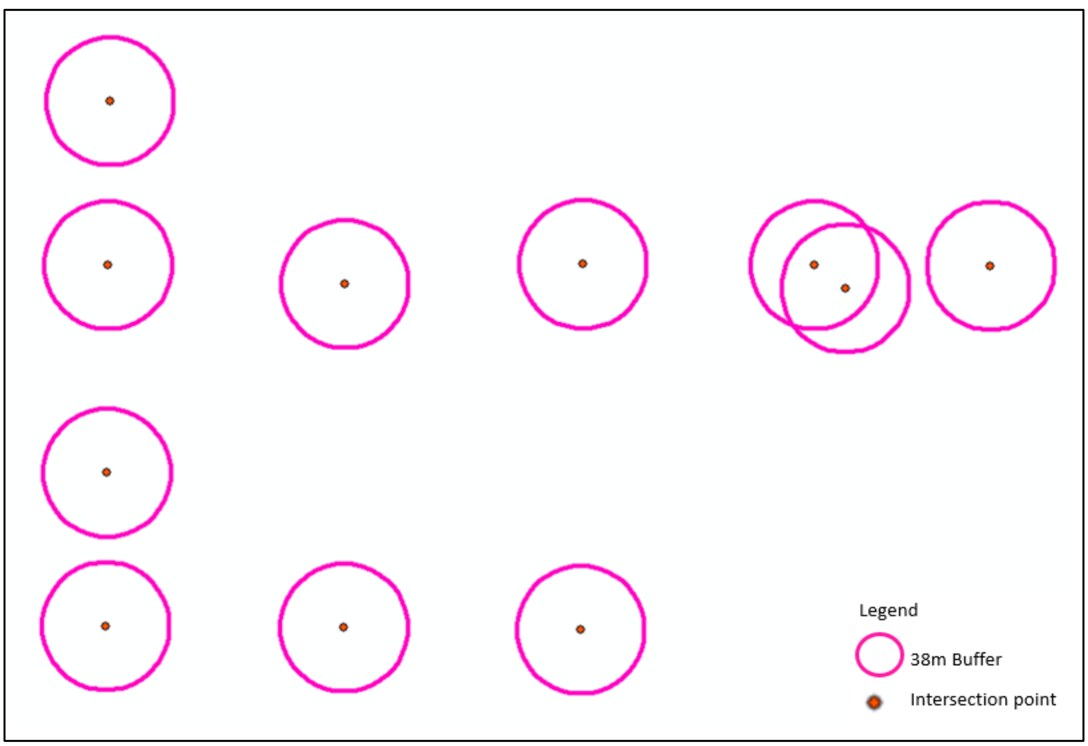

Figure 4.4: Buffered intersection points after data cleaning

4.3 Data Processing for Exploratory Spatial Data Analysis

Data processing for ESDA was necessary to obtain the aggregated count of MV traffic crashes at intersections within the analysis unit. Geographic Information system was used to identify locations in the study area where the highest count of MV traffic crashes at intersections were concentrated. ArcGIS Pro 2.4 is the software tool used for the GIS Analysis.

The study area was selected from the Des Moines area using the select features tool in GIS. Roads in the study area were obtained using the intersect tool to intersect the study area and Des Moines roads. Points were placed where roads met each other in the study area, at intersections using the point feature creation tool. The Point features created at road intersections (Intersection points) served as the centroid to create Thiessen polygons.

The number of intersection points in the study area was 1,891. Figure 4.5 shows the study area within the City of Des Moines with its roads and intersection points.

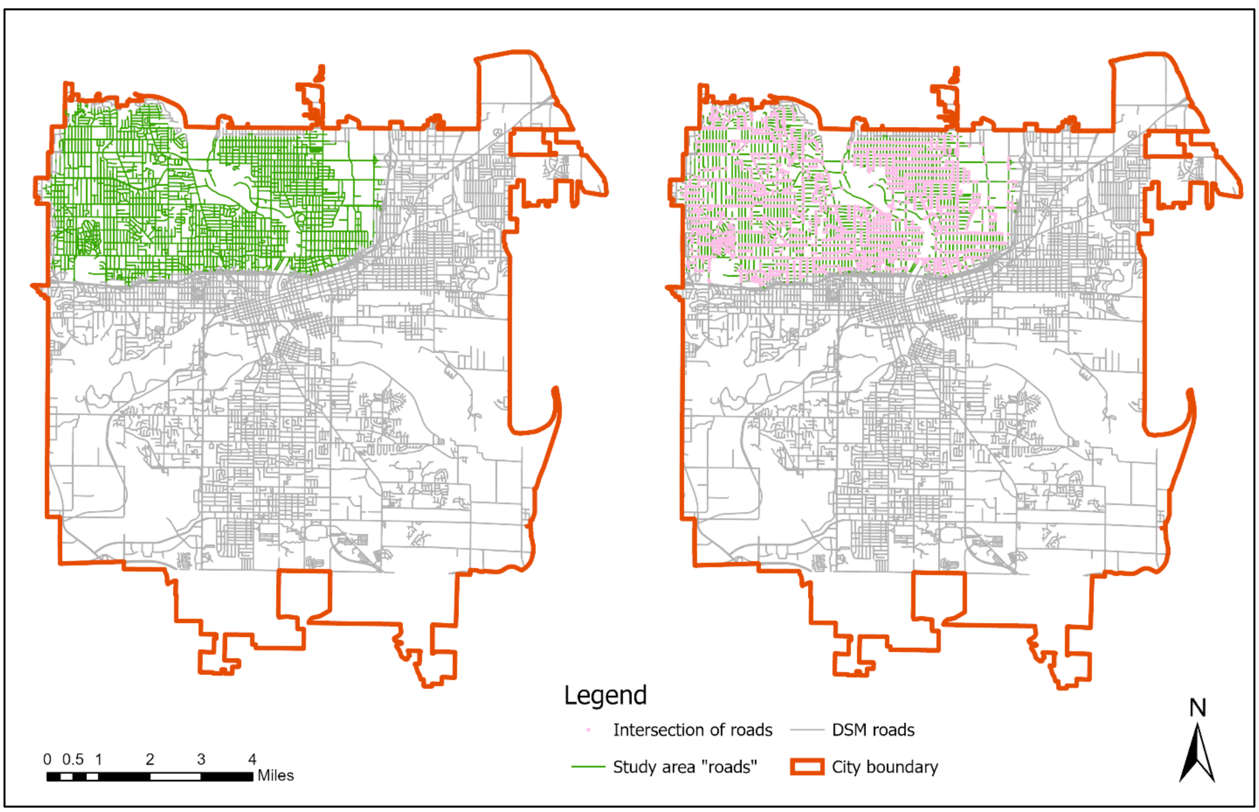

Figure 4.5: City of Des Moines roads and intersection points

Thiessen polygons were created to serve as the unit of analysis in the exploratory spatial data analysis process to identify high and low clusters of intersection related MV traffic crashes. The Thiessen polygons generate neighboring relationships for exploratory spatial data analysis; hence the intersection point acting as a centroid will be the neighbor of another.

The perpendicular bisectors of the lines between all the intersection points are the foundation for Thiessen polygons creation. Figure 4.6 displays the roads and the intersection points surrounded by Thiessen polygons.

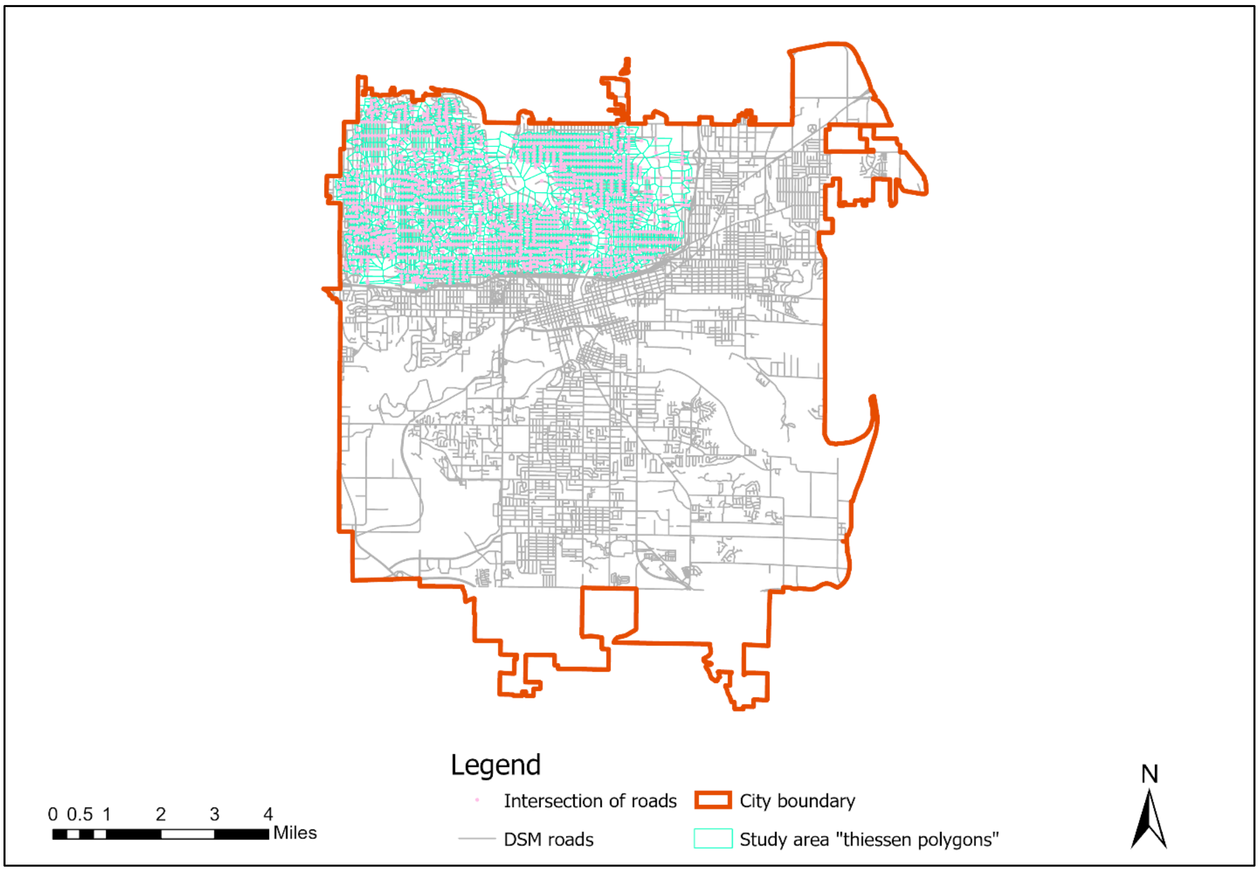

Figure 4.6: Thiessen polygons with roads and intersection points

Buffering was employed in this study to define a specified radius or width around the intersection points since the study focusses on intersection related MV crashes. Buffering was implemented using the buffer tool. A buffer tool is a geoprocessing tool that creates polygons that surrounds features within a specified distance. A 38-meter buffer was therefore created around each intersection point to capture intersection related traffic crashes. Crashes are considered as intersection related if they are within 250ft from the center of any approach intersection (Li, 2018). Figure 4.7 shows all intersection points in the study area within a 38-meter buffer, and figure 4.8 shows a zoomed-in view of the buffered intersection point.

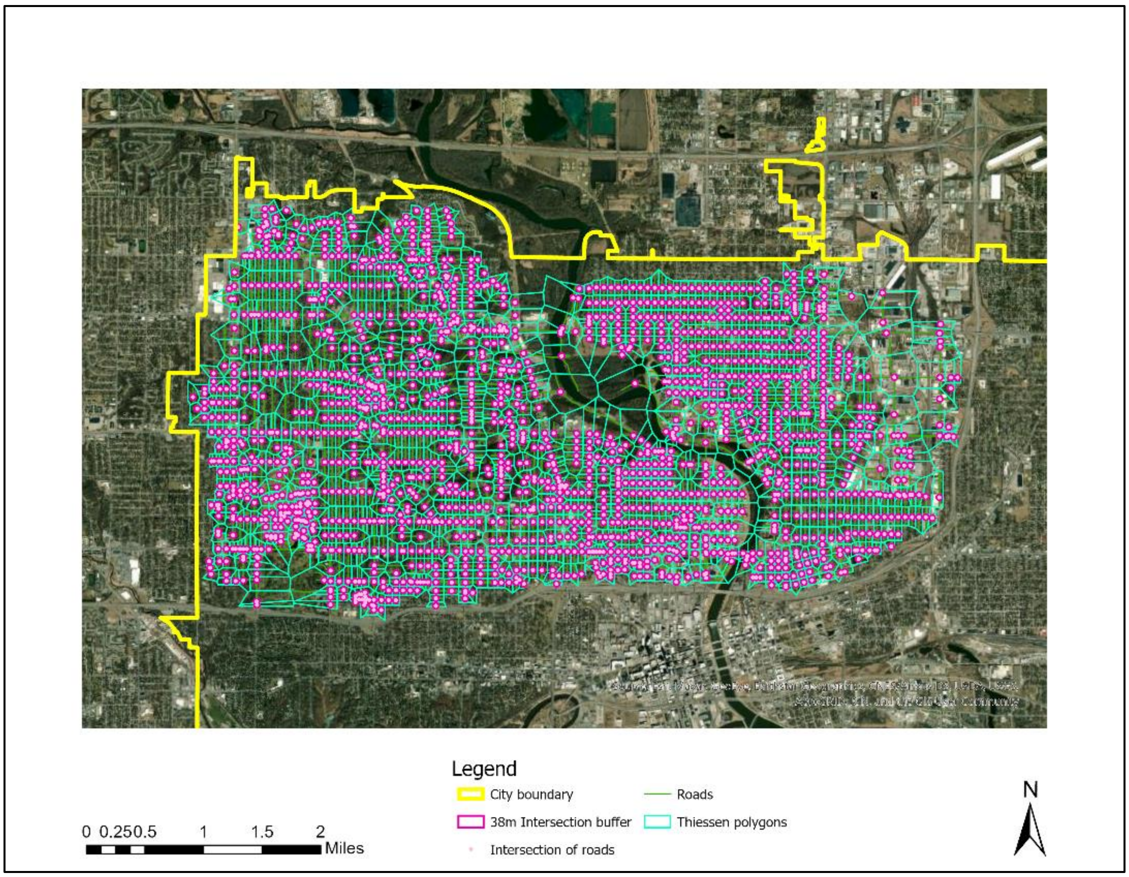

Figure 4.7: Buffered intersections with Thiessen polygons

Figure 4.8: Zoomed in buffered intersections with Thiessen polygons

The 7-year motor vehicle crash data (2013-2019) collected from the IDOT was loaded into the study area, as shown in figure 4.9. Location data (x and y coordinates) in the crash data enabled geocoding to be executed in GIS. Geocoding is transforming a location description, such as a pair of coordinates, to a location on the earth's surface. Geocoding displays the crash data spatially and plays a crucial role in detecting motor vehicle crash patterns.

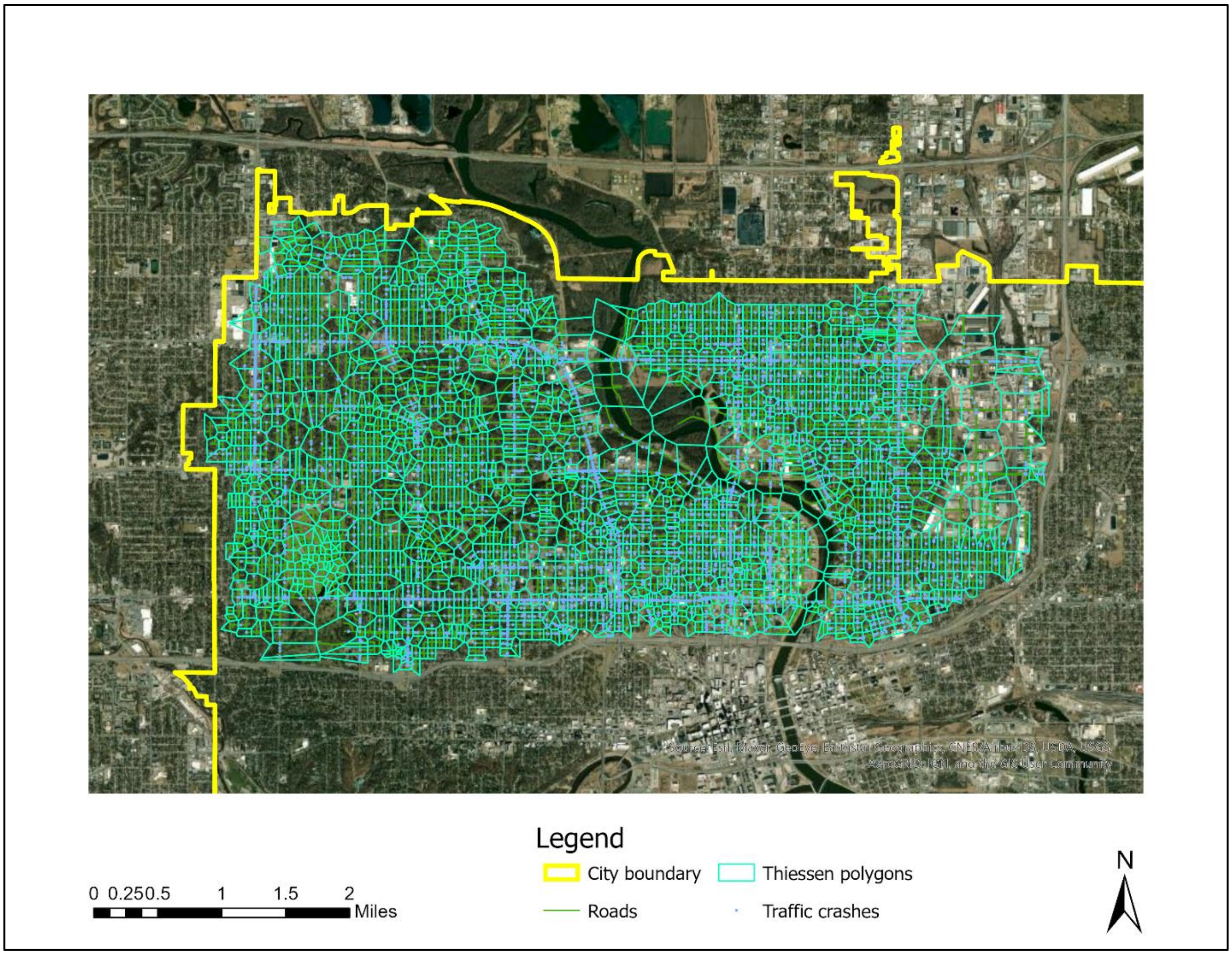

Figure 4.9: MV traffic crashes (2013-2019)

The spatial join feature was employed to obtain the count of point features in a polygon feature, with the MV traffic crashes acting as point features and the buffered intersection points acting as polygon features. Hence, the count of crashes around each buffered intersection point was obtained by employing the GIS's spatial join tool. The spatial join tool pairs rows from the join feature to the target feature using their relative spatial locations (Esri, 2020). The count of traffic crashes in each 38m buffer polygon is shown in figure 4.10.

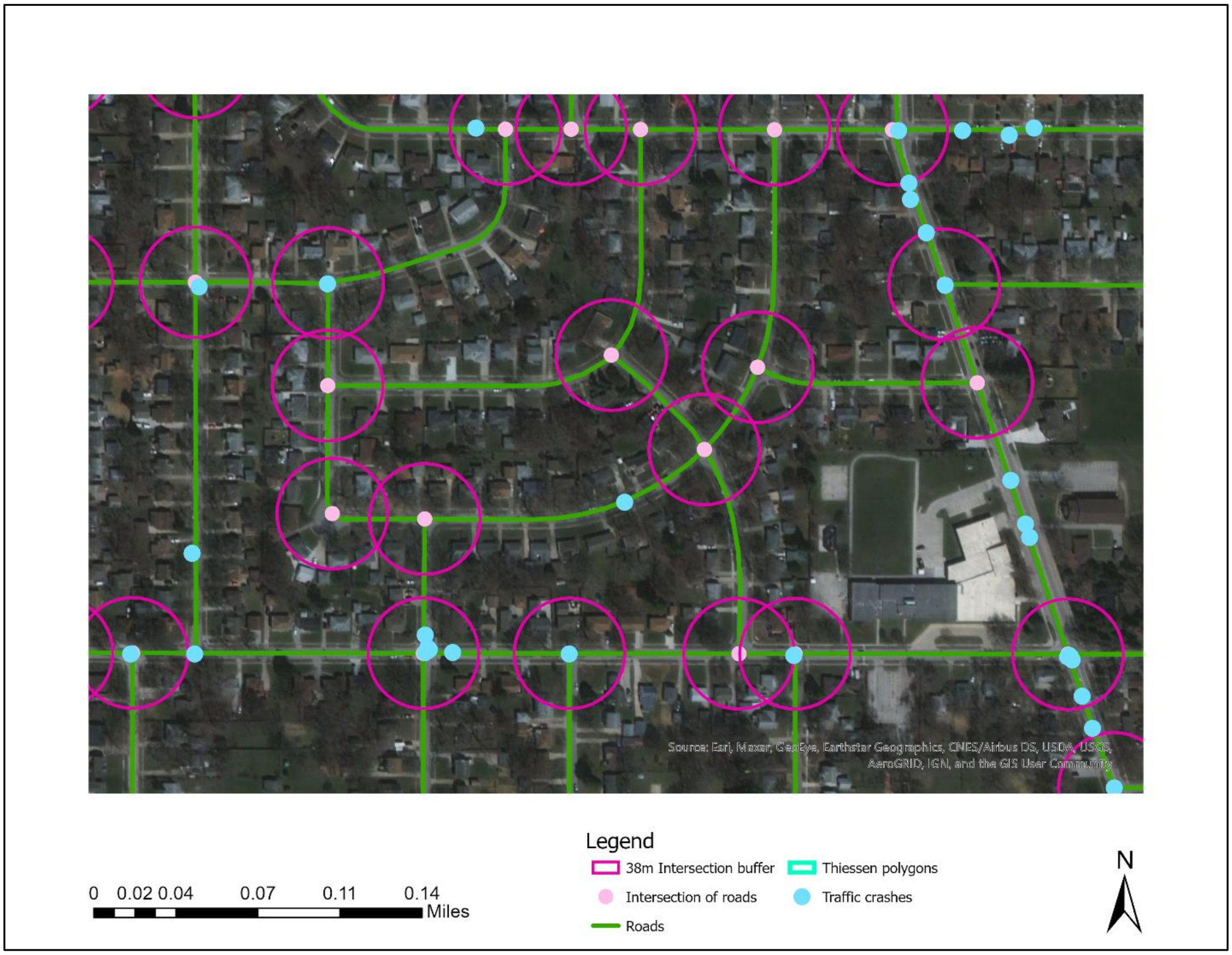

Figure 4.10: Buffered intersections with MV traffic crashes

A choropleth map was produced for the count of MV traffic crashes, as shown in figure 4.11. The choropleth map shows the variability of the count of MV traffic crashes, which is represented by different shaded patterns. The choropleth map alone of MV crashes does not show patterns of clustering. To display the patterns of clustering of MV crashes at hotpots intersections, exploratory spatial data analysis (ESDA) is used. ESDA employs specific techniques to visually explore MV crash data in space to identify and analyze spatial patterns.

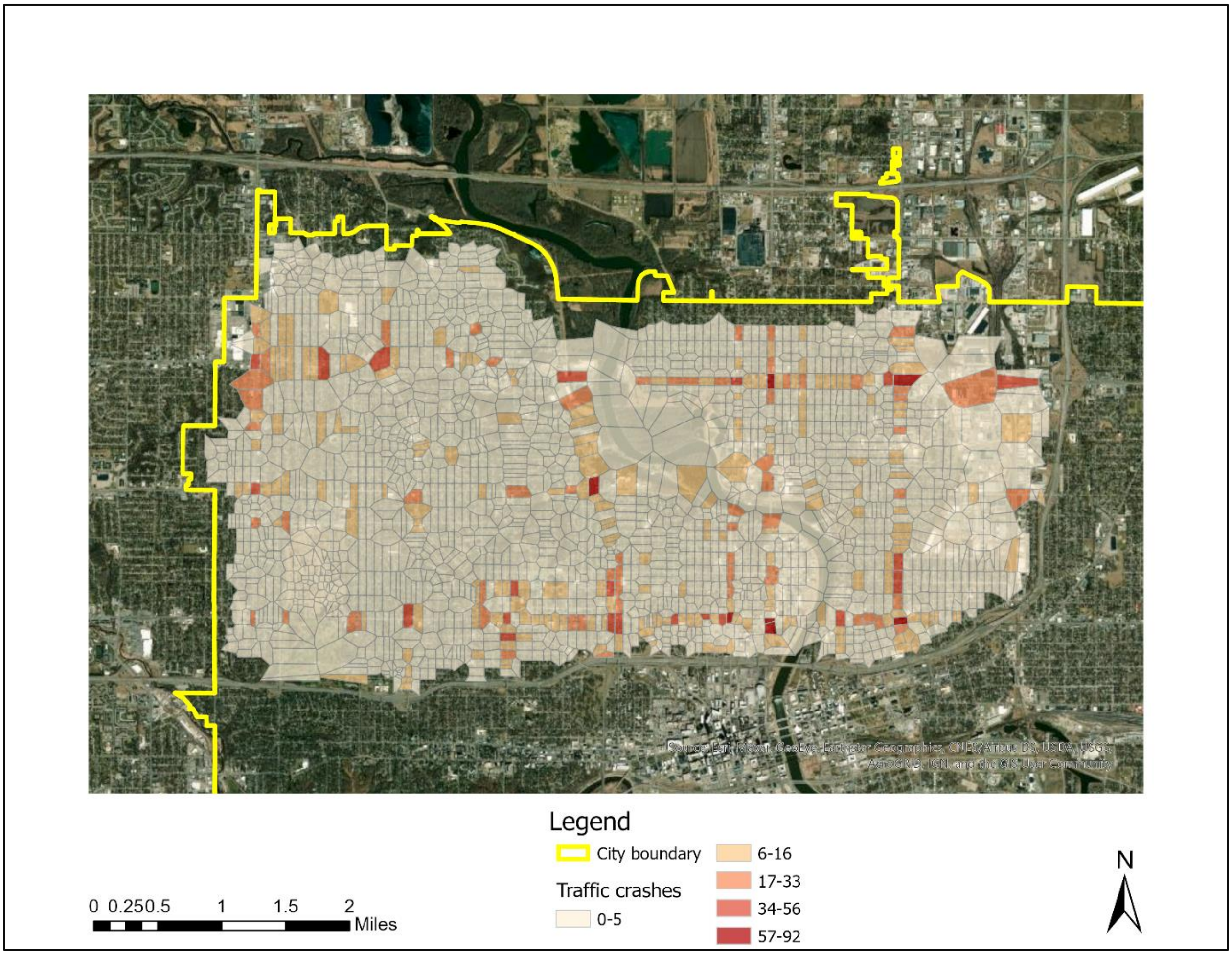

Figure 4.11: Choropleth map of MV intersection traffic crashes (2013-2019)

4.4 Exploratory Spatial Data Analysis (ESDA)

An essential concept of ESDA is spatial autocorrelation, which is the coincidence of value similarity and locational similarity. Spatial autocorrelation is a concept that allows an understanding of weather location matters in the spatial distribution of a variable (Anselin,1995). In this study, it was used to examine if high values of traffic crashes are surrounded by high values of traffic crashes (high clusters) and if low values of traffic crashes are surrounded by low values of traffic crashes low clusters). Spatial autocorrelation would reveal whether the traffic crashes are randomly distributed in the study area or not. Geoda 1.14.0 is the software tool used for ESDA in this study.

This study uses two types of statistics to measure spatial autocorrelation; one measure is at the global level, and the other measure is at the local level. Moran's I statistic is used at the global level to indicate the clustering of MV traffic crashes. At the local level, the Local Indicators of Spatial Association (LISA) helps understand clusters of MV traffic crashes. To perform ESDA, spatial weight matrices' construction is necessary to estimate Global Moran's I and LISA.

4.4.1 Spatial Weight Matrix

Spatial weight is an important element used to test spatial autocorrelation statistics and allows for cross-sectional spatial dependence analysis. The spatial arrangement of the data is influenced by the spatial weight matrix, which also captures location similarity (Anselin, 1995). In this study, two spatial weight matrices were constructed to determine the output's sturdiness, the queen contiguity matrix, and the K- nearest neighbor matrix.

The queen contiguity weight matrix is a kind of binary contiguity matrix defined by a common border and a common vertex (Anselin, 2020). The characteristic of the contiguity relationship is denoted as

$$w = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are contiguous} \\ 0 & \text{if } i \text{ and } j \text{ are not contiguous} \end{cases}$$

The K-nearest neighbor matrix is a kind of distance-based weight in which the K-nearest observations have the same number of neighbors for all observations regardless of distance (Anselin, 2018). The K-nearest neighbors' matrix is denoted as

$$w_{ij}^{*}(k) = 0 \quad \text{if } i = j$$

$$w_{ij}^{*}(k) = 1 \quad \text{if } d_{ij} \leq d_i(k) \text{ and } w_{ij}(k) = w_{ij}^{*}(k) \Big/ \sum_j w_{ij}^{*}(k)$$

$$w_{ij}^{*}(k) = 0 \quad if\ d_{ij} > \ d_i(k)$$

Where di (k) is a critical cut-off distance for each Thiessen polygon i; di (k) is the kth order smallest distance between Thiessen polygons i and j such that each Thiessen polygon i has exactly k neighbors. For this study, the number of neighbors (K) is 7, and this is determined based on the highest frequency of the histogram of the queen contiguity matrix, as shown in figure 4.12.

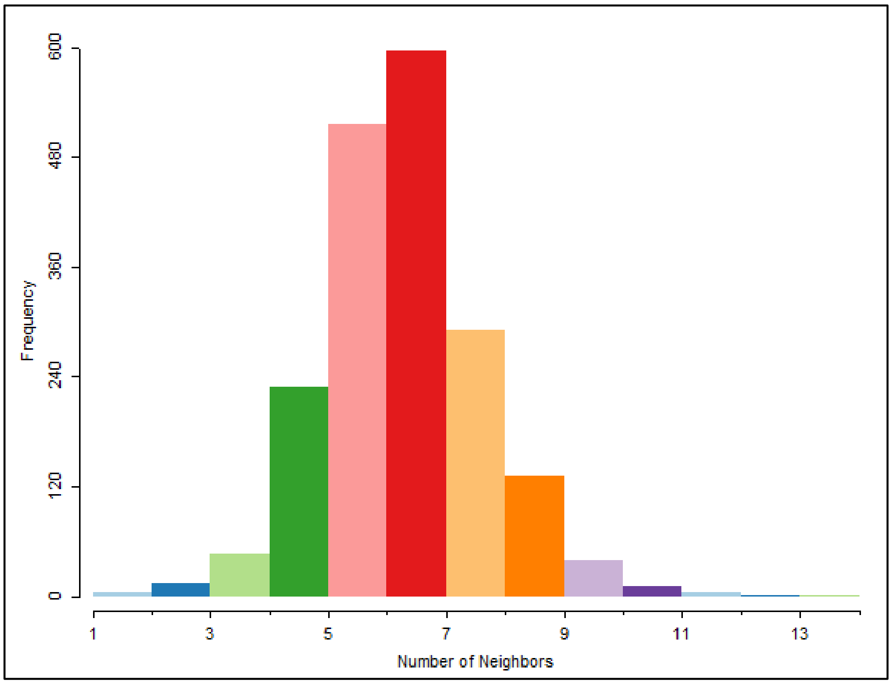

Figure 4.12: Histogram of queen contiguity matrix

4.4.2 Moran's I

Moran's I statistic measures global spatial autocorrelation, and it is a cross-product statistic between a variable and its spatial lag, with the variable expressed in deviations from its mean. The spatial lag is the weighted average of the neighboring values (Anselin, 2018). The Moran's I measures whether the pattern of MV traffic crashes at intersections is clustered, random, or dispersed.

The Moran's *I* statistic for spatial autocorrelation denoted as:

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} z_i z_j}{S_0 \quad \sum_{i=1}^{n} z_i^2}$$

Where zi is the deviation of an attribute for feature i from its mean (xi – X), wi,j is the special weight between feature i and j, n is equal to the total number of features, and S0 is the aggregate of all the spatial weights.

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}$$

The Moran's scatter plot is a visual tool that enables an assessment of the similarity of an observed value to its neighboring values. It is a plot with an x-axis that shows the original variable: the count of MV intersection traffic crashes variable, and a y-axis that shows the spatially lagged count of MV intersection traffic crashes variable. The slope of the linear fit to the scatter plot equals Moran's I.

Moran's scatter plot displays spatial autocorrelation classification into four categories from the four quadrants of the plot. These four categories are high high (HH), low low (LL), high low (HL), and low high (LH). Similar values surrounded by similar values, which is positive spatial autocorrelation, correspond to HH and LL, which is shown in the upper-right quadrant and the lower-left quadrant, respectively. HH refers to high values of MV traffic crashes surrounded by high values in neighboring locations. LL refers to low values of MV traffic crashes surrounded by low values in neighboring locations. Dissimilar values surrounded by dissimilar values, that is, negative spatial autocorrelation, correspond to HL and LH, shown in the lower-right quadrant and the upper-left quadrant. HL defines high values of MV traffic crashes surrounded by low values in neighboring locations. LH represents low values of MV traffic crashes surrounded by high values in neighboring locations.

Moran's Values range between 1 and -1 representing exact positive spatial autocorrelation and exact negative spatial autocorrelation, respectively. However, it cannot be interpreted directly, but it is well understood within the context of a null hypothesis. Hypothesis testing is conducted when working with spatial autocorrelation. It is a method that employs sample evidence to determine if a specific hypothesis about a population is true or false. The null hypothesis is a hypothesis that is tested during hypothesis testing, and it is the assumption of the non-existence of spatial autocorrelation and the existence of spatial randomness. If the null hypothesis is false, then the hypothesis is assumed to be true, implying the existence of an alternative hypothesis. The alternative hypothesis deals with the presence of spatial autocorrelation, which can be positive or negative (Anselin, 2018).

The Moran's scatter plot for both the queen and the K-nearest neighbor is shown in figure 4.13 and figure 4.14, respectively. As can be observed from the Moran's I scatterplot of both the queen contiguity matrix and the K-nearest neighbor matrix, clustering of most observations occurs in the upper right of the quadrant than the other three quadrants. The Moran's I values for both the queen contiguity matrix and the K-nearest neighbor matrix are positive; hence there is a positive spatial autocorrelation. The existence of clustering is confirmed. Therefore, the null hypothesis can be rejected since there is a pattern, which means that spatial randomness is absent; accordingly, there is a spatial structure.

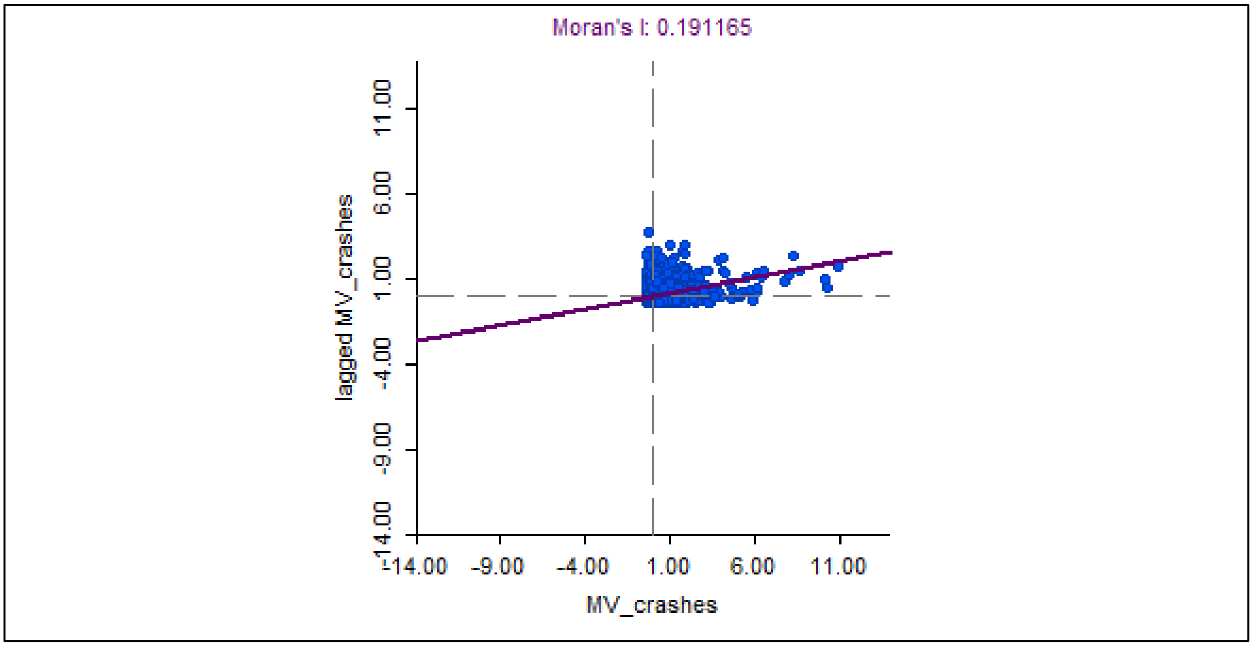

Figure 4.13: Scatter plot (Queen contiguity matrix)

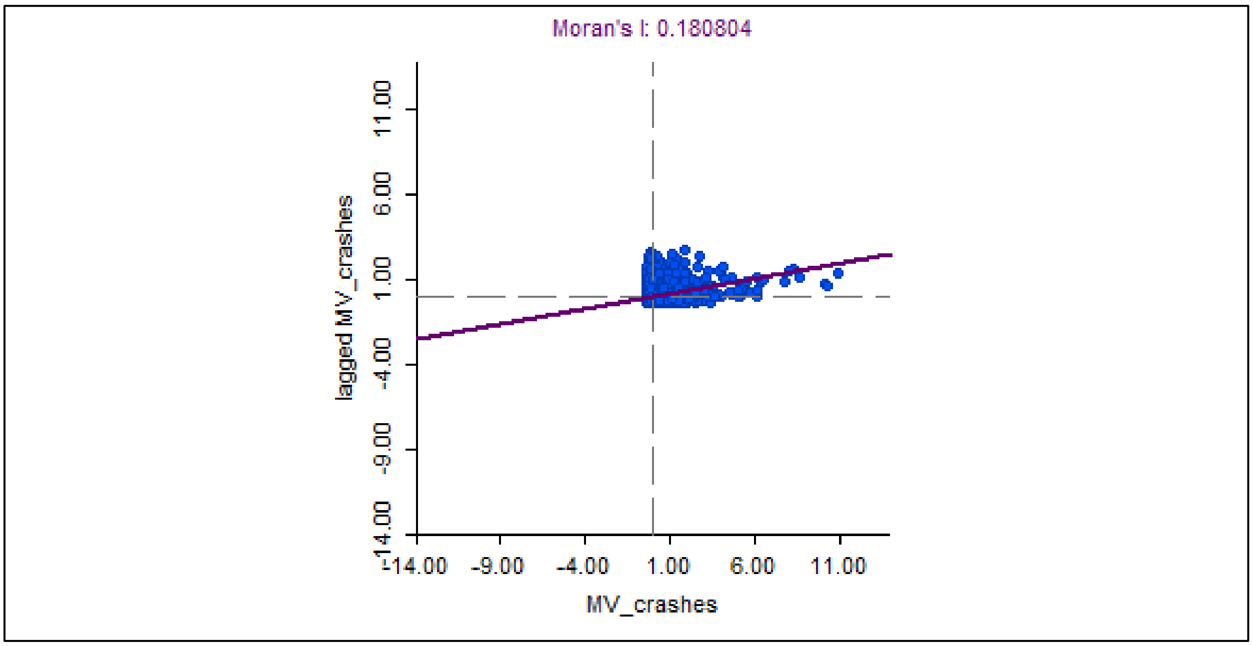

Figure 4.14: Scatter plot (K-nearest neighbor matrix)

A permutation was run to assess the significance of the test statistic. The permutation run generates an empirical reference distribution. It allows quantification of how extreme the observed statistics are, relative to what its distribution would be under the null hypothesis of spatial randomness (Anselin, 2018). The summary statistics from the reference distribution were pseudo value (p-value), the number of permutations runs, Moran's I value, the theoretically expected value, mean, standard deviation, and z-value. The reference distribution of the queen contiguity matrix and K-nearest neighbor is shown in figure 4.15.

The number of permutations that is run affects the stability of the p-value and influences the degree of significance (Anselin, 2018). In this study, Global Moran's I for queen contiguity matrix and the K-nearest neighbor matrix were constructed with a 1% level of significance and a significance test of 999 permutations, resulting in stable results.

The P-value is denoted as

$$p = \frac{R+1}{M+1}$$

Where R=permuted data sets and M= permutation number

The Z-value is denoted as

$$z_I = \frac{I - E[I]}{\sqrt{V[I]}}$$

Where:

$$E[I] = -1/(n-1)$$

$$V[I] = E[I^2] - E[I]^2$$

n= number of features

I = Moran's I value

A positive z-value and a statistically significant p-value represent high and or low values' spatial clustering. A negative z-value, combined with a statistically significant p-value represents a spatial dispersion of high and low values (Esri, 2020).

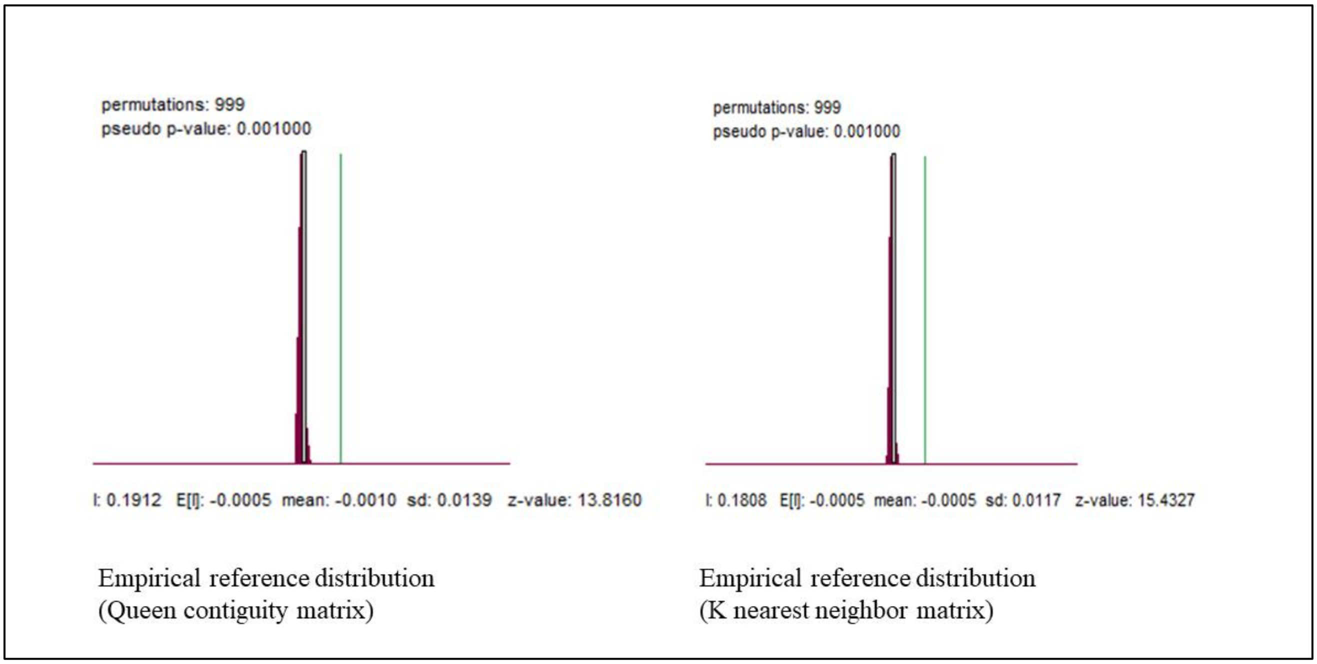

Figure 4.15: Empirical reference distribution

The Global Moran's I provide information on clustering but is limited in providing information on spatial clusters or outliers (Anselin, 1995); therefore, local indicators of spatial association was implemented.

4.4.3 Local Indicators of Spatial Association (LISA)

Local Indicators of Spatial Association (LISA) was employed to provide information about clusters of traffic crashes in the study area by identifying exact locations that contribute to a global pattern. Unlike Global Moran's scatter plot, LISA shows the significance and indicates the existence or non-existence of significant spatial clusters or outliers for each location. LISA allows for the decomposition of global indicators, such as Moran's I, into each observation (Anselin, 1995). LISA is mathematically defined as

$$I_i = \frac{(x_i - \mu)}{m_0}\sum_j w_{ij}(x_j - \mu) \text{ and } m_0 = \frac{(x_i - \mu)^2}{n}$$

Where xi is the observation in the Thiessen polygon, i is the mean of the observations across the Thiessen polygons. The summation over j is such that only neighboring values of j are included.

Cluster and significance maps were generated after performing LISA, which was used to identify MV crashes clusters. The cluster and significance map provides information on the number of HH, LL, HL, and LH Thiessen polygons. The cluster maps for the queen contiguity matrix and K-nearest neighbor matrix are shown in figures 4.16 and 4.18, respectively. The significance maps for the queen contiguity matrix and K-nearest neighbor matrix can be seen in figures 4.17 and 4.19, respectively. More observations were not significant in the queen contiguity matrix than the K-nearest neighbor. The K-nearest neighbor had more clusters for HH and LL than the Queen contiguity matrix.

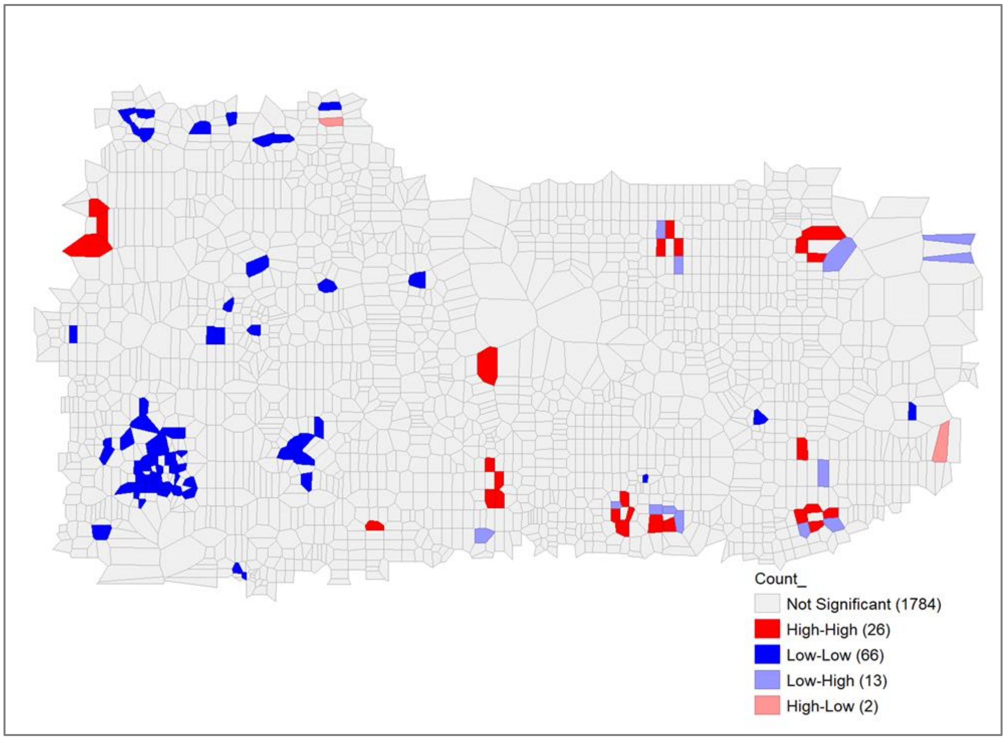

Figure 4.16: Cluster map (Queen contiguity matrix)

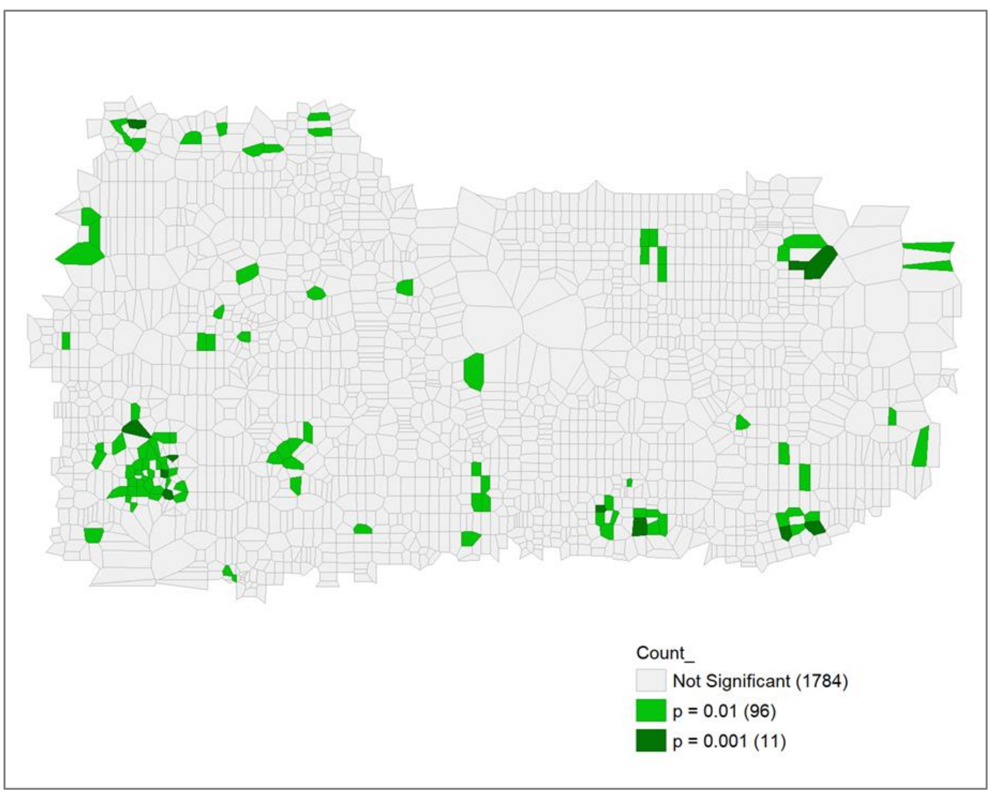

Figure 4.17: Significance map (Queen contiguity matrix)

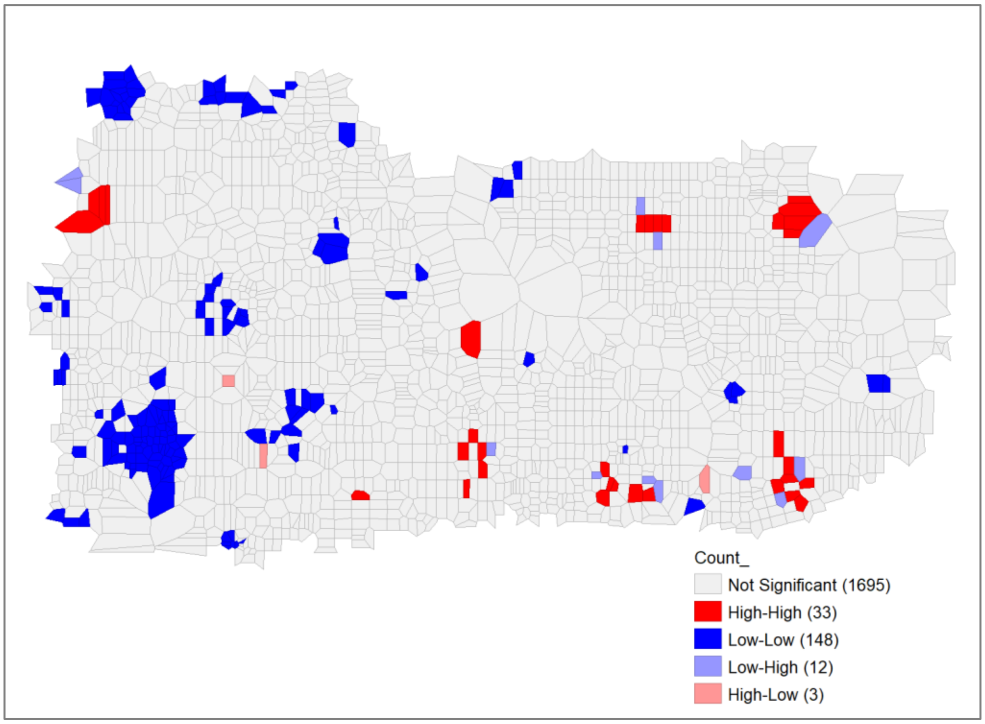

Figure 4.18: Cluster map (K-nearest neighbor matrix)

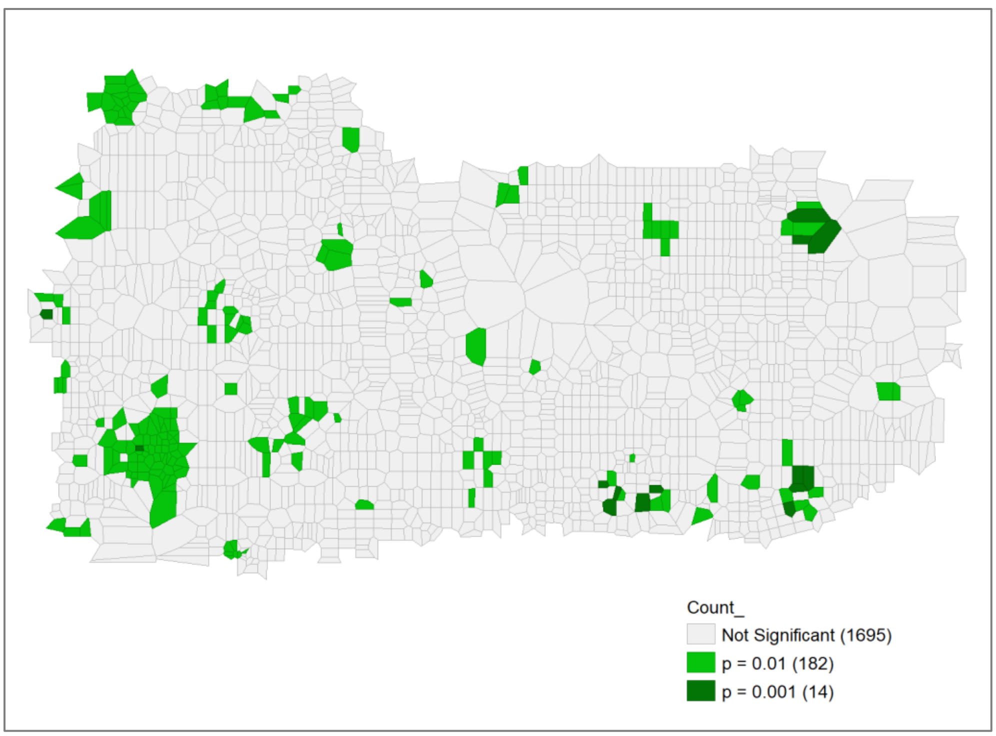

Figure 4.19: Significance map (K-nearest neighbor matrix)

4.5 Questionnaire Design

A GSV questionnaire was developed from built environment variables identified from previous research, as shown in table 4.1, that led to having a positive or negative relationship with MV traffic crashes. The questionnaire was mostly close-ended questions and dichotomous; that is, they either demanded a yes or no response as to whether a built environment variable was identified in google street view or not. Goggle street view images were observed from 2013-2019 on all the intersection legs, the northbound leg, southbound leg, eastbound leg, westbound leg., etc. The duration was chosen to match the years of the motor vehicle crashes analyzed. The GSV questionnaire is shown in appendix A. A total of 12 built environment variables were examined from the GSV images.

Table 4.1: Built environment variables and traffic crashes

Author	**Study Area**	**Built Environment Variables**
Ouyang & Ilir Bejler (2014)	Miami–Dade County, Florida	Mixed land use (+) Number of bus stops (+)
Wolf & Bratton (2006)	United States	Trees (+) Guardrails (+) Utility Poles (+)
Zahabi et al. (2011)	Montreal, Canada	Signalized intersections (+) 10m of park (+) Arterials (+) Local Streets (-)
Gladhill & Monsere (2012)	Portland, Oregon	Schools (-) Transit stops (+) Business density (+) 4-leg intersection (+)
Dezman, et al. (2016)	Baltimore, Maryland	High density center (+)
Soltani & Askari (2017)	Shiraz, Iran	Arterial roadway (+) Multi-lane roadway (+) Urban activity centers (+)

Table 4.1 continued

Author	Study Area	Built Environment Variables
Pulugurtha et al. (2013)	Charlotte, North Carolina	Single family residential (-) Institutional area (+) Industrial area (+) Research district area (+)
Kim & Yamashita (2002)	Honolulu, Hawaii	Commercial land uses (+) Visitor lodging (+) Manufacturing (+) Retail (+) Office space (+)
Kim et al. (2006)	Honolulu, Hawaii	Parks (+) Commercial activities (+) Schools (+)
Huang et al. (2018)	Detroit, Michigan	Commercial land use (+) Four-way intersections (+)
Kaygisiz et al. (2017)	Eskisehir, Turkey	Number of transit stops (+) Mixed land uses (+)

4.6 Statistical Analysis

This study employs multiple linear regression, stepwise multiple linear regression, and the Tobit model to analyze the relationship between MV traffic crashes and the built environment. Multiple regression is an extension of simple linear regression and requires at least two independent variables: binominal, ordinal, or interval or ratio level variables. It is used to predict a variable's value based on the value of two or more other variables. The variable to be predicted is called the dependent variable, and for this study, the MV intersection crash rate acts as the dependent variable. The independent variables influence the dependent variable; therefore, the built environment variables were the independent variables. A multiple linear regression model with k independent variables x_1, x_2... x_k and a dependent variable y, can be written as:

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon.$

Where $\beta_0, \beta_1, \beta_2 \ldots \beta_k$ are the regression coefficients and ϵ is the residual term.

The sample size criterion is that regression analysis requires at least 20 cases per independent variable in the study. The multiple linear regression analysis's key assumptions include linearity, multivariate normality, no multicollinearity, and homoscedasticity. Linearity refers to the existence of a linear relationship between the outcome variable and the independent variables, and scatterplots show whether there is a linear or curvilinear relationship. Multivariate Normality assumes that the residuals are normally distributed in the multiple regression. No multicollinearity is the assumption that the independent variables are not highly correlated with each other. Variance Inflation Factor (VIF) values are used to test this assumption. Homoscedasticity assumes that the variance of error terms is similar across the independent variables' values (Statistics solutions, 2020).

The stepwise multiple linear regression entails an iterative process of adding and removing predictors to obtain the significant variables in the dataset that results in the best fit model, which is a model that best expresses the relationship between the dependent and independent variables and lowers prediction error (Frost et al., 2020).

Tobin (1958) proposed an approach called the "Tobit" model, also known as the censored regression model, to obtain consistent estimates when the dependent variable is censored. In statistical analyses, if the dependent variable is censored for a significant number of the observations, conventional regression methods result in biased parameter estimates. Therefore, the Tobit model estimates linear relationships between variables when either left, or right-censoring exists in the dependent variable. The left censoring, referred to as censoring from below, censor's data at a low threshold. The right censoring is also known as censoring from above and censor's data at a high threshold.

The Tobit model is defined as

$$y_i^* = x_i' \beta + \varepsilon_i$$

The subscript i = 1, . . . , N represents the observations, y_i^* is a latent variable, x_i' is a vector of explanatory variables, β is a vector of unknown parameters, and ε_i represents the disturbance term.

$$y_i = \begin{cases} a & if\ y_i^* \leq a \\ y_i^* & if\ a < y_i^* < b \\ b & if\ y_i^* \geq b \end{cases}$$

Where a is the lower limit, and b is the upper limit of the dependent variable. If a = −∞ or b = ∞, the dependent variable is not left-censored or right-censored, respectively (Henningsen, 2010).

CHAPTER 5. RESULTS

5.1 ESDA Results

The results from the empirical reference distribution are displayed in table 5.1. The z-values and p-values indicate whether to reject the null hypothesis or not. A very small p-value means there is a probability that the observed spatial pattern is due to random processes is unlikely; hence the null hypothesis can be rejected. The p-values for both the K-nearest neighbor matrix and the queen contiguity matrix was small; thus, the null hypothesis was rejected, and an alternative hypothesis was assumed.

Table 5.1: Statistical results

Variable	**K-Nearest Neighbor Matrix**			**Queen Contiguity Matrix**		
	Moran’s I	P-value	Z-value	Moran's I	P-value	Z-value
MV intersection traffic crashes	0.1912	0.0001	14.3884	0.181	0.0001	15.4327

The cluster and significance map results of the K-nearest neighbors’ matrix and the queen contiguity matrix were saved as attributes in the Thiessen polygon shapefile. The Thiessen polygon shapefile was imported into a GIS environment to identify high and low clusters of MV intersection traffic crashes. A query definition was employed to identify HH and LL Thiessen polygons. The query definition included the two spatial weight matrices to narrow down clusters significant in both spatial weight matrices.

The query definition for was HH Thiessen polygon QLISA_CL=1 AND QLISA_P <= 0.01 AND NLISA_CL=1 AND NLISA_P <=0.01 and that for LL Thiessen polygons was QLISA_CL=2 AND QLISA_P <= 0.01 AND NLISA_CL=2 AND NLISA_P <=0.01. QLISA_CL and QLISA_P represent cluster map results and significance map results of the queen contiguity matrix, respectively. NLISA_CL and NLISA_P represent the cluster map results and significance map results of the K-nearest neighbor matrix, respectively. Figure 5.1 and figure 5.2 shows the Thiessen polygons with high and low clusters of MV intersection traffic crashes.

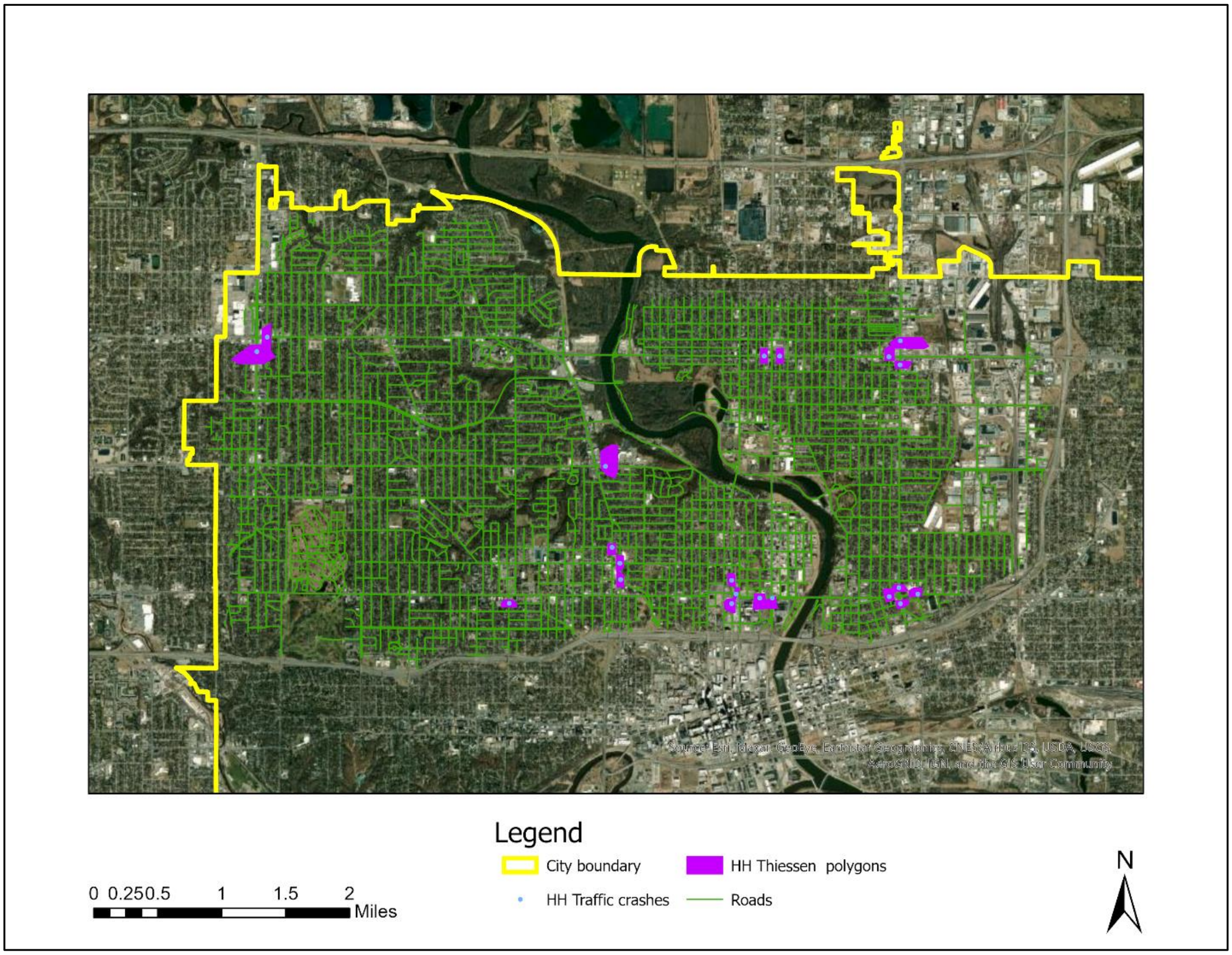

Figure 5.1: HH Thiessen polygons with HH MV traffic crashes

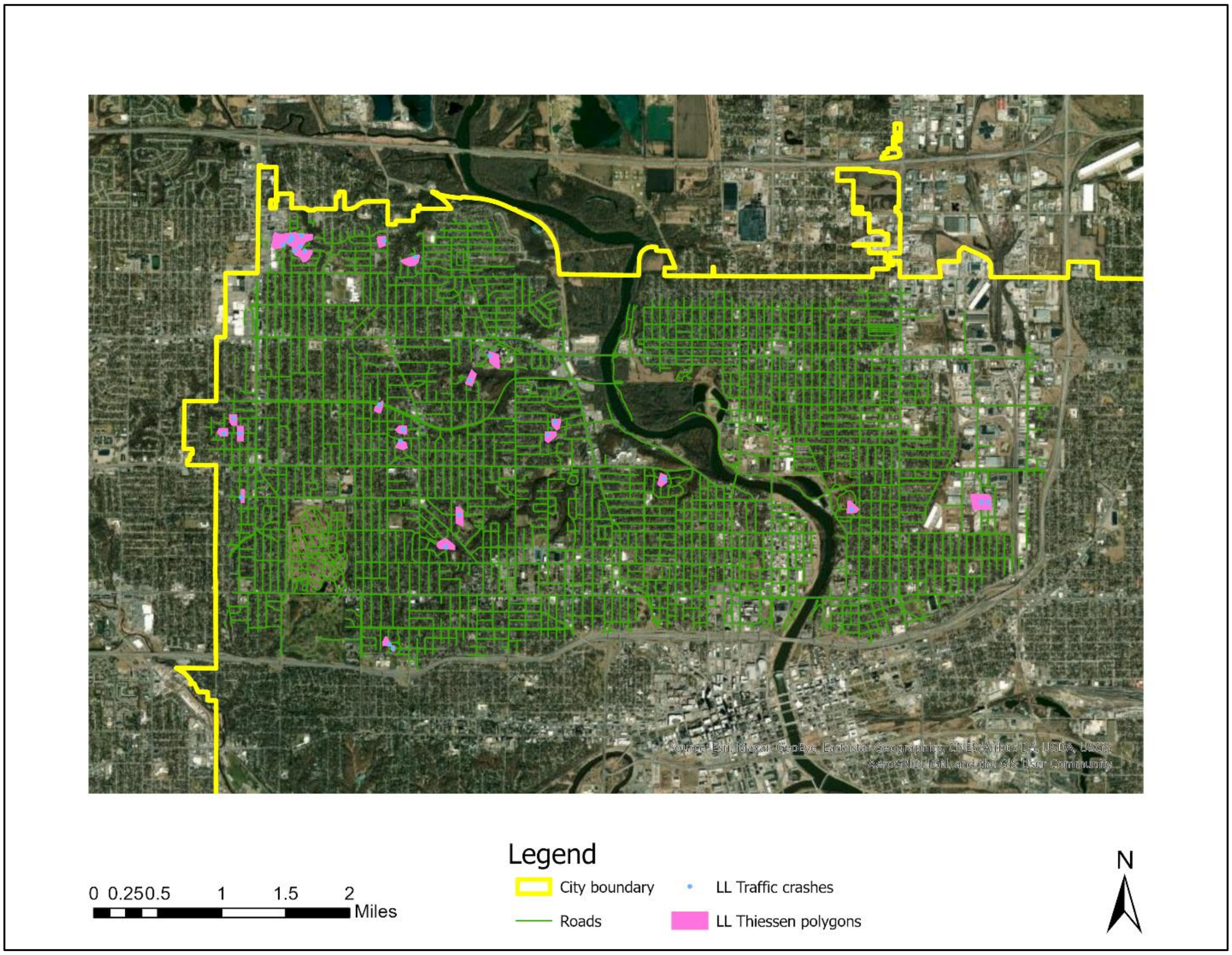

Figure 5.2: LL Thiessen polygons with LL MV traffic crashes

The neighborhoods where high and low clusters of MV traffic crashes were distributed were also obtained. The study area neighborhood shapefile was loaded in the GIS environment. The select tool was used to select the neighborhoods with HH Thiessen polygons with HH traffic crashes, as shown in figure 5.3 and LL Thiessen polygons with LL traffic crashes were as shown in figure 5.4. The selected neighborhoods were then exported as a shapefile using the export selected features tool.

The neighborhoods with HH clusters of MV traffic crashes at intersections were Merle Hay, Drake, Highland Park, Prospect Park, King Irving, Cheatom Park, Capitol Park, and Martin Luther Jr. Parkway. The neighborhoods with LL clusters of MV traffic crashes at intersections were Waveland Park, Lower Beaver, Beaverdale, Prospect Park, Chautauqua Park, Merle Hay, and Union Park.

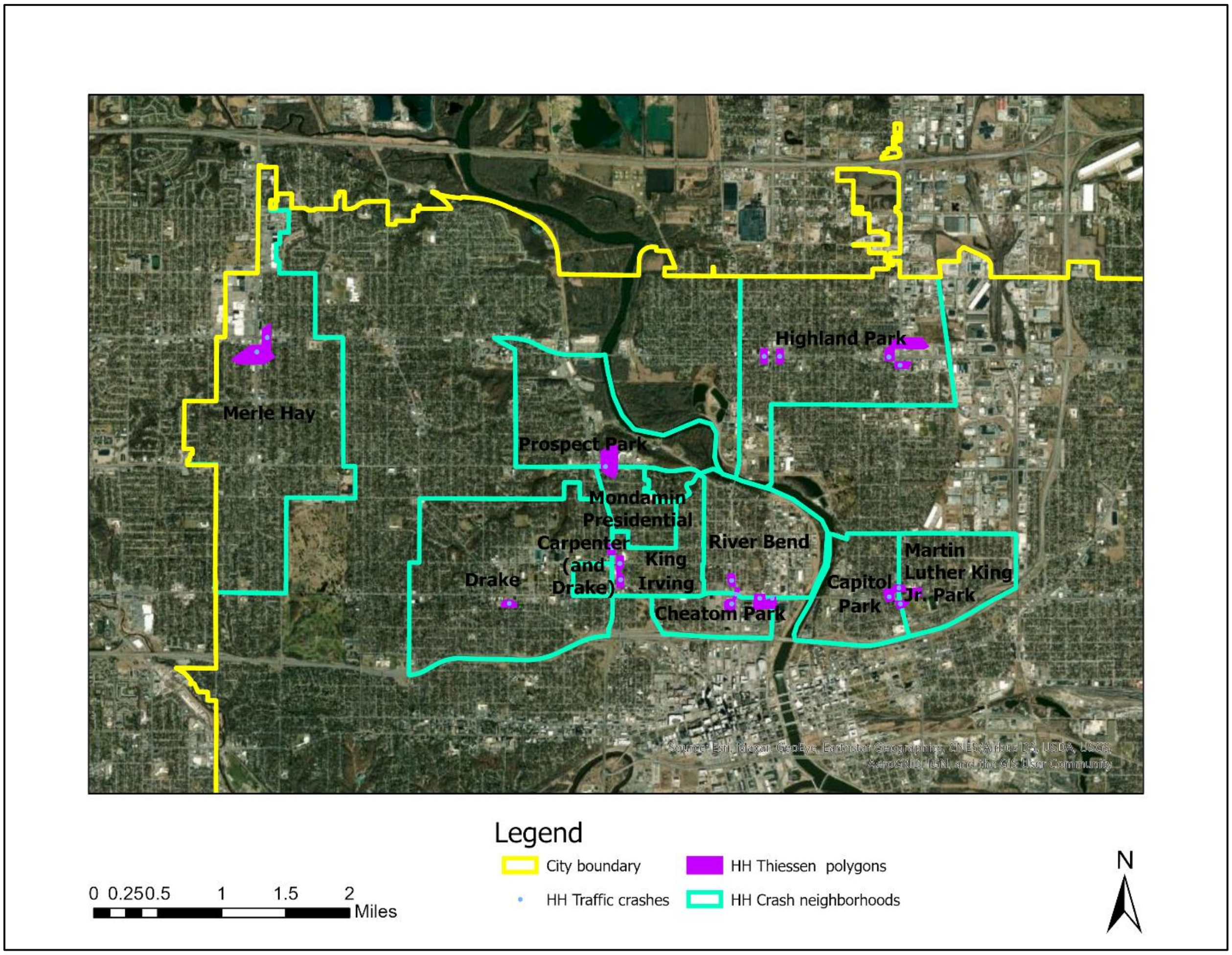

Figure 5.3: Neighborhoods with HH MV traffic crashes

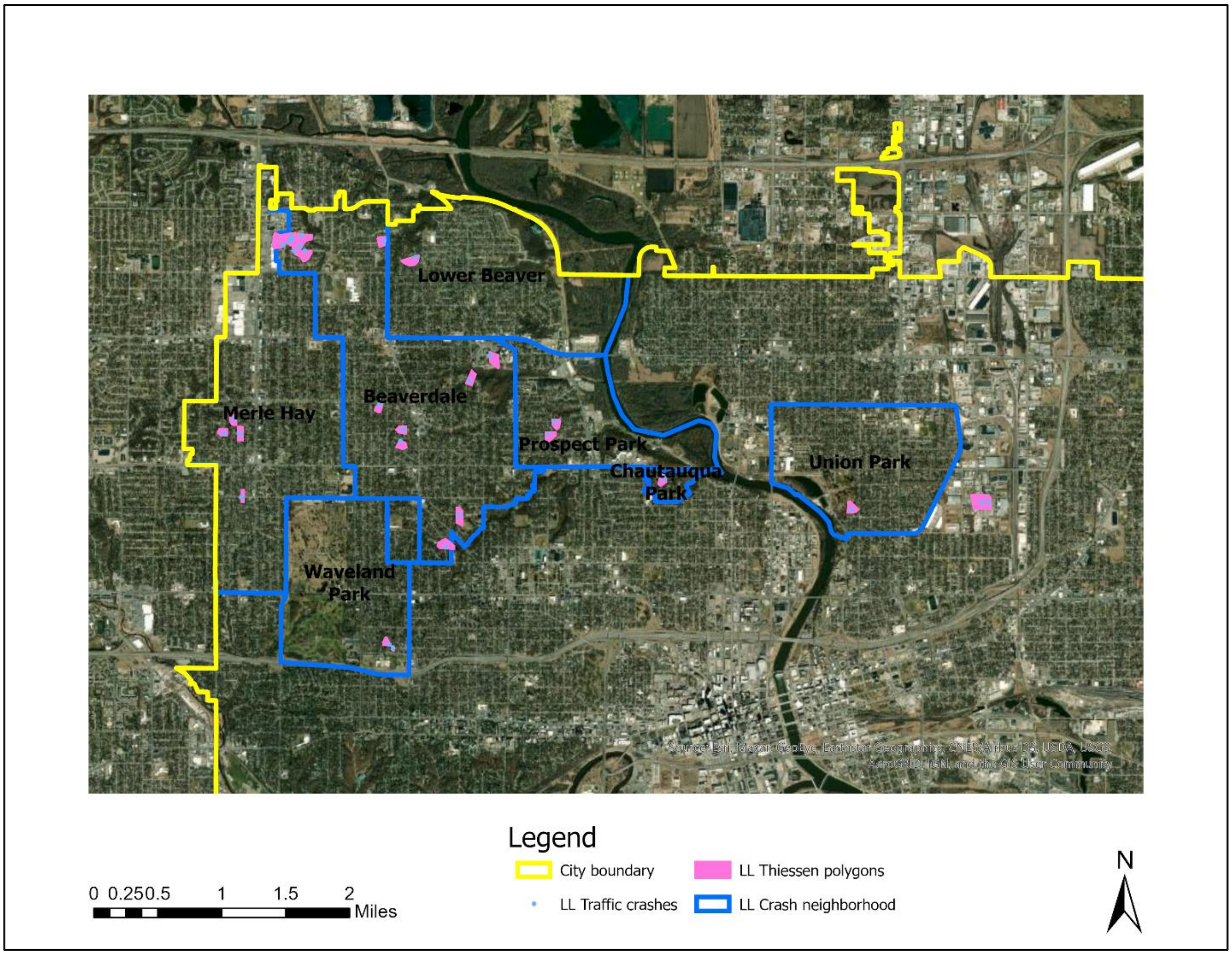

Figure 5.4: Neighborhoods with LL MV traffic crashes

5.2 Level of Poverty and MV Traffic Crashes at Intersections

LL and HH Thiessen polygons were displayed on a choropleth map of household poverty in the study area, as shown in figure 5.5. The LL Thiessen polygons are in the maps' light shaded areas, representing wealthy neighborhoods, while the HH Thiessen polygons are in the dark shaded areas, which represent low-income neighborhoods. Even though figure 5.5 shows a high occurrence of MV intersection traffic crashes occurs in low-income neighborhoods, it does provide information on the extent of association.

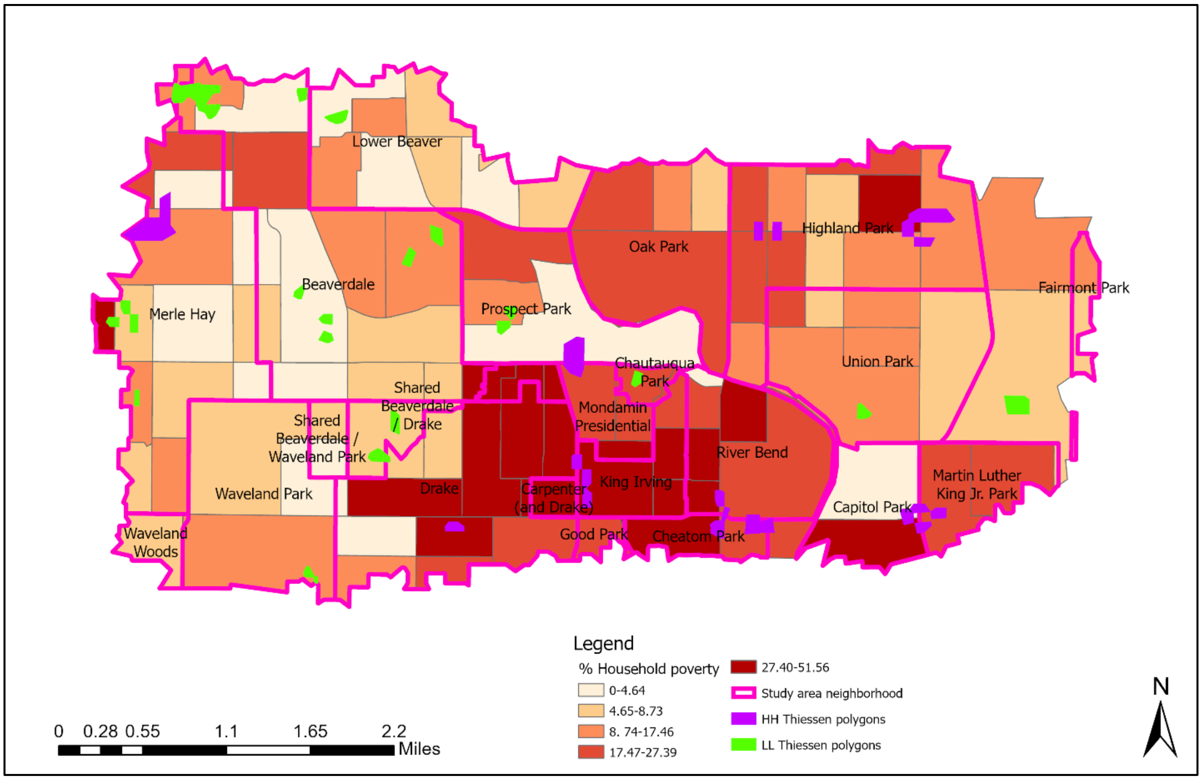

Figure 5.5: Choropleth map of household poverty with HH and LL Thiessen polygons

A raster map of household poverty shown in figure 5.6 was created to obtain poverty values in high and low clusters of MV intersection traffic crash locations for inclusion as an independent variable for statistical analysis. The raster map was created using the features to raster spatial analyst tool in GIS, and the poverty values were obtained using the extract values to points tool in GIS.

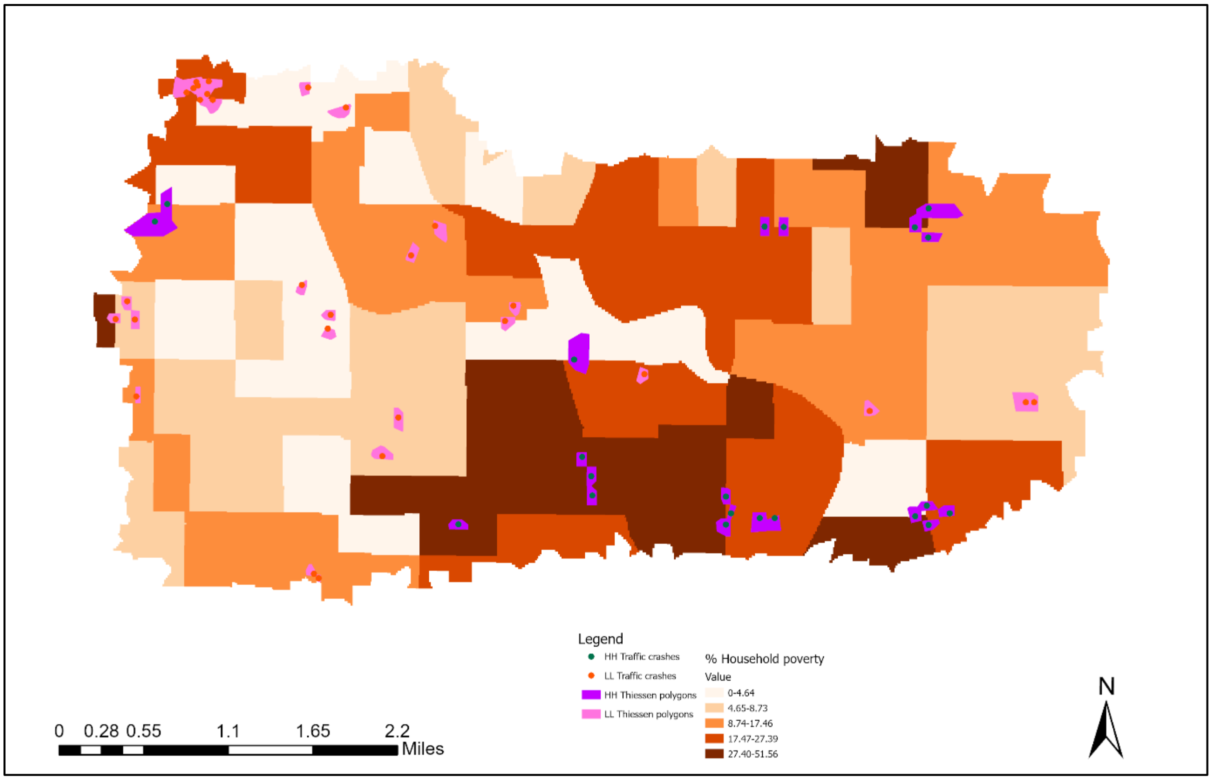

Figure 5.6: Household poverty raster map

5.3 Google Street View Results

The empirical studies of Google Street View conclude that virtual fieldwork can be replaced with in situ fieldwork; thus, the built environment variables in both high and low clusters of MV traffic crash intersections were examined using GSV. The questionnaire was conducted for 40 intersection points, as shown in figure 5.7 and within a radius of 40m from the intersection. Built environment variables assessed included commercial/institutional land use, single-family residential, schools, parks, signage, street lanes, on-street parking, bus stops, trees, crosswalks, priority intersections, and signalized intersections.

Built environment variables were observed in GSV from 2013 through 2019, which is the same as years of traffic crash data. A standard questionnaire was employed to examine the built environment variables observed on all intersection legs, i.e., the northbound leg, the southbound leg, the eastbound leg, the westbound leg, etc. The intersections legs in the study area included 3-

legged, 4-legged, and 5-legged intersections controlled for in the regression modeling. Built environment variables on each intersection leg were given a score between 0 and 1 on each intersection leg based on its existence or non-existence. The built environment data were processed in MS Excel to have a maximum of 1 if it existed on all intersection.

The Polk county assessor's website was utilized to identify the classification of properties near the intersections where necessary. The website enables the entering of the street address where the property is, and information is provided on whether the property is residential, commercial, or industrial. Figure 5.8 shows an identification of a property using the Polk county assessor's website.

The GSV questionnaire results were prepared using Microsoft Excel with a yes response eliciting a 1 and a no response prompting a 0. The results of all intersection legs at an intersection were converted to a percentage, and line diagrams were produced in MS Excel to compare HH and LL crash intersections. The results were then imported into statistical software (R studio 4.0.1) for further analysis.

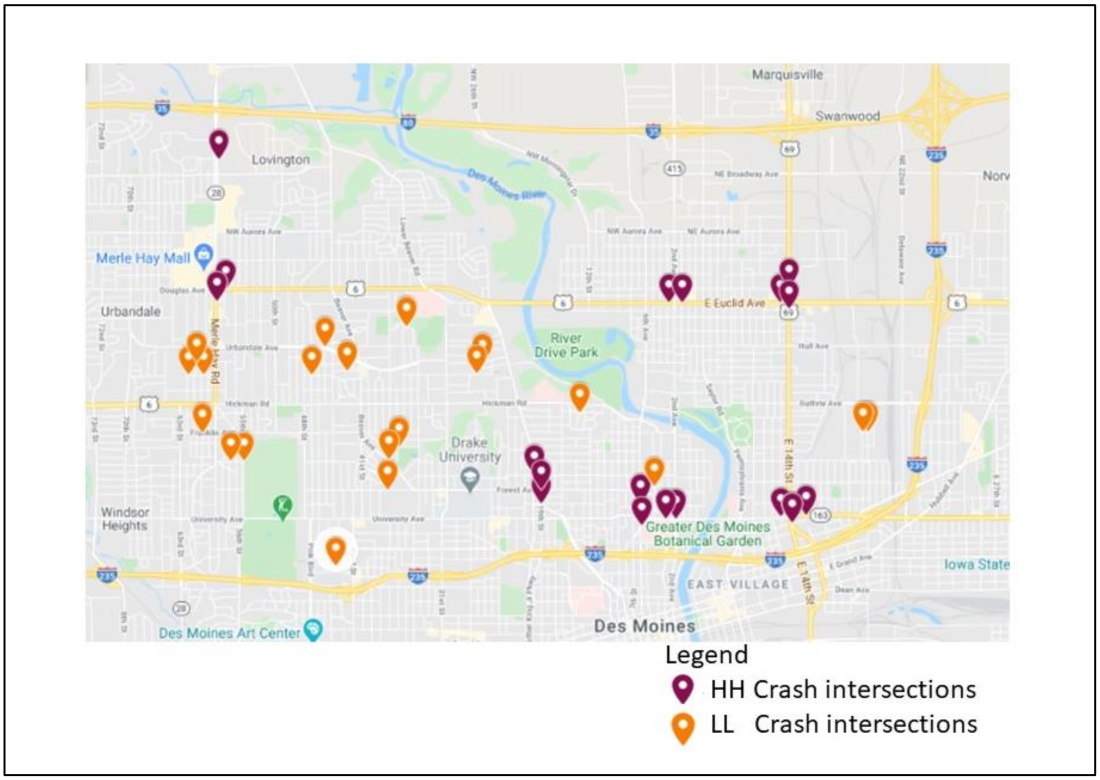

Figure 5.7: A google map of HH and LL MV crash intersections

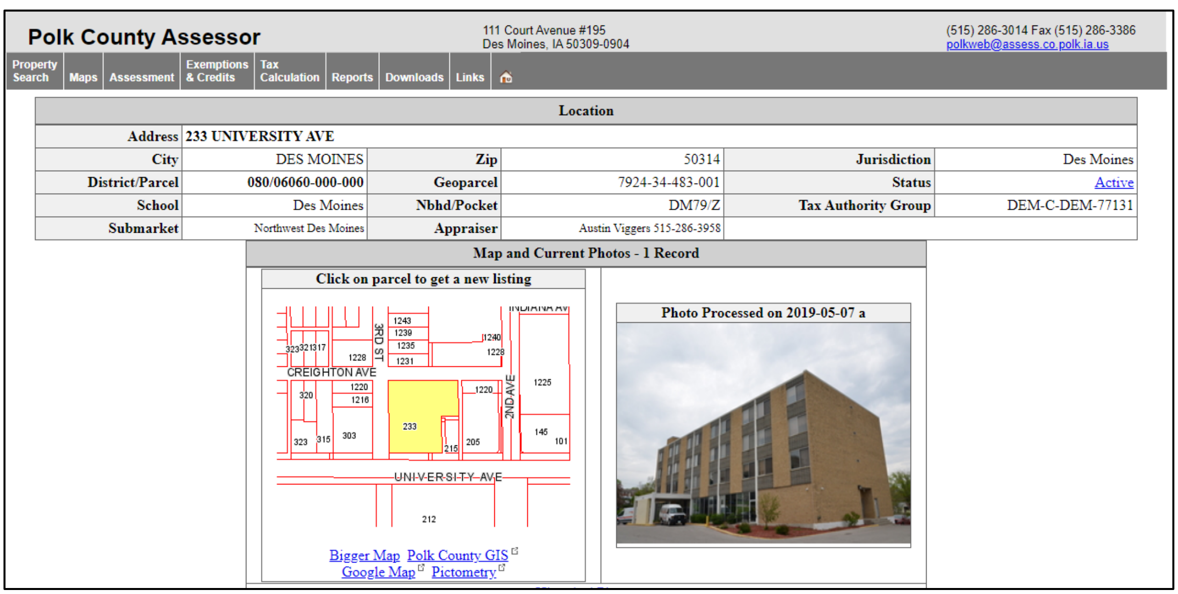

Figure 5.8: Polk county assessor

The summary statistics of the variables are shown in table 5.2 and was obtained using MS Excel. The built environment variables have a minimum of 0 indicating the absence of variable, and a maximum of 1 indicating its presence on all intersection legs.

Table 5.2: Summary statistics of variables

Variable Name	Mean	Median	S. D	Min	Max
MV intersection crash rate	60.87	44.56	62.54	0	214.98
Household Poverty	16.76	13.79	12.58	1.43	51.56
Commercial/institutional land uses	0.4	0.38	0.42	0	1
Single family residential	0.7	1	0.46	0	1
Signalized intersections	0.1	0	0.3	0	1
Sidewalks	0.73	1	0.45	0	1
Crosswalks	0.08	0	0.27	0	1
Priority Intersections	0.58	1	0.5	0	1
Schools	0.03	0	0.11	0	0.5
Parks	0.05	0	0.15	0	0.5
Signage	0.31	0	0.43	0	1
On-street parking	0.38	0	0.47	0	1
Bus stops	0.24	0	0.47	0	1
Trees	0.78	1	0.42	0	1

In this study, commercial land uses were defined as land designated for profit-generating activities, and institutional land uses in this study were defined as land designated for public buildings except for schools and parks. Commercial land uses included office complexes, shopping malls, service stations, restaurants, retail space, visitor lodging, laundry, etc. Institutional land uses included hospitals, orphanages, clinics, churches, senior living facilities, government organizations, non-governmental organizations, and community centers. Figure 5.9 shows a plot of the commercial/institutional land uses in the HH and LL crash intersections. It can be observed that commercial/institutional land uses are more present in HH crash intersections than LL crash intersections.

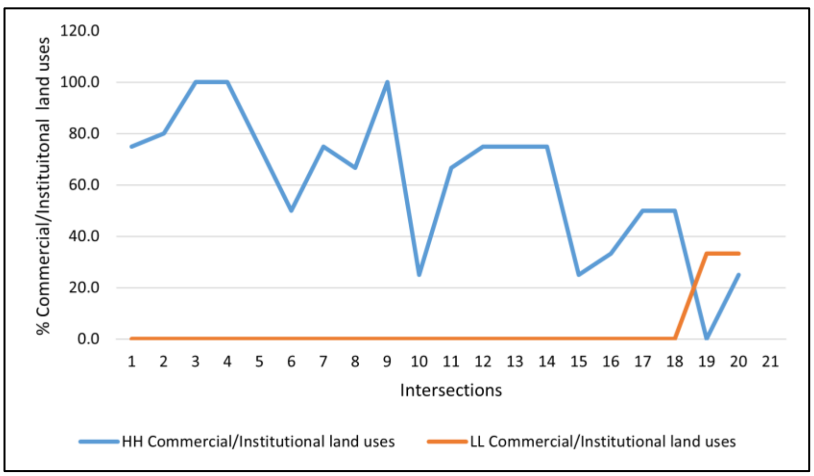

Figure 5.9: Line graph of commercial/institutional land use

Single family residential was defined in this study as a residential building that caters to a single family. Figure 5.10 shows a plot of single family residential in the HH and LL crash intersections. It indicates that LL crash intersections are mainly in single family residential only areas compared to HH crash intersections.

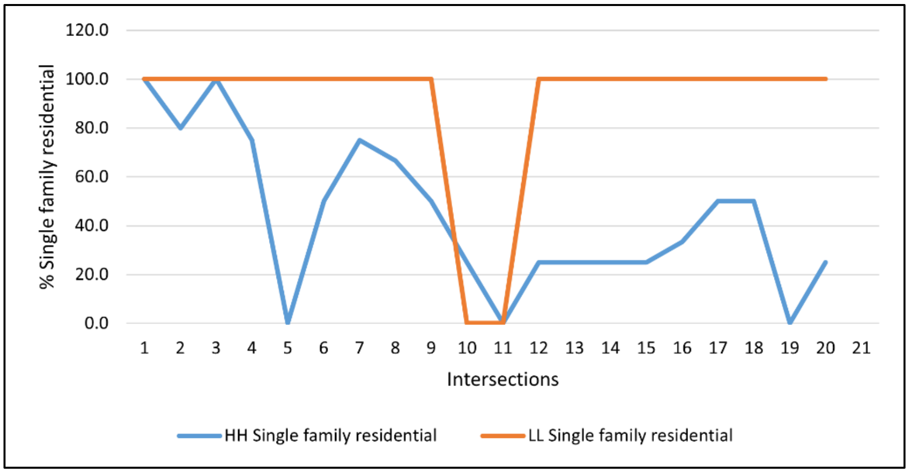

Figure 5.10: Line graph of single family residential

Schools were defined as an educational facility that provides learning spaces for educating students. In this study, Schools were a stand-alone independent variable, not inclusive of institutional land use because it generates trips at specific times of the day. Figure 5.11 shows a plot of schools in the HH and LL crash intersections, and it shows that the schools in the study area were mainly the HH crash intersections.

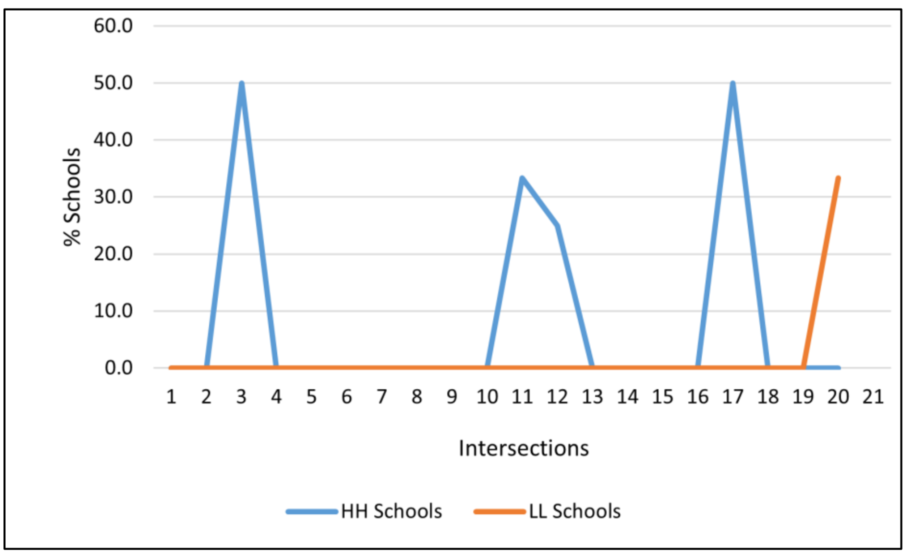

Figure 5.11: Line graph of schools

Signage in this study was defined as the use of signs and symbols to relay a message. Signage was mostly observed at commercial/institutional land uses. Figure 5.12 shows that the signage was mainly in the HH crash intersections than LL crash intersections.

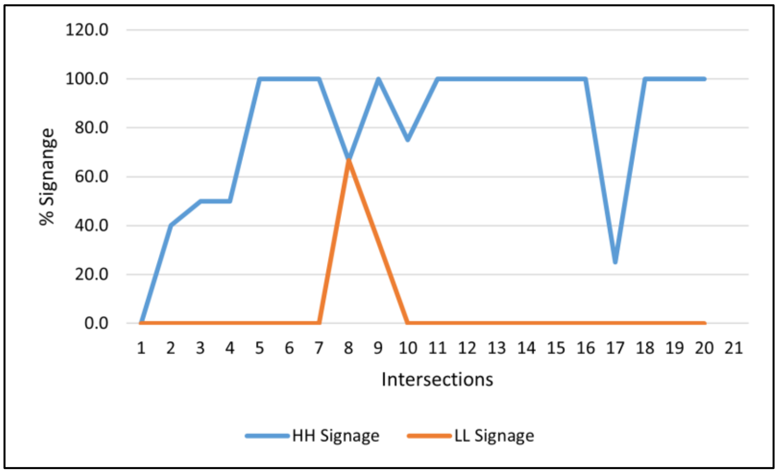

Figure 5.12: Line graph of signage

On-street parking was defined as a parking space on the edge of the street. Figure 5.13 shows that the on-street parking was mainly in the LL crash intersections than HH crash intersections. On-street parking was more prevalent in LL crash areas where land use was mostly monotonous than HH crash areas.

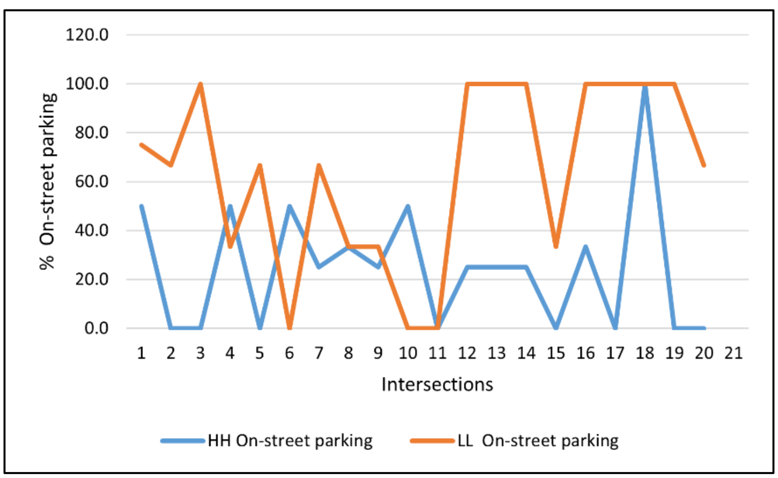

Figure 5.13: Line graph of on-street parking

Bus stops were defined as a location designated for passengers to get on or off a bus. Figure 5.14 shows that the bus stops were mainly in the HH crash intersections than LL crash intersections. The GSV questionnaire results revealed bus stops were mostly observed in areas with a mix of land use, i.e., commercial, institutional, and or single family residential.

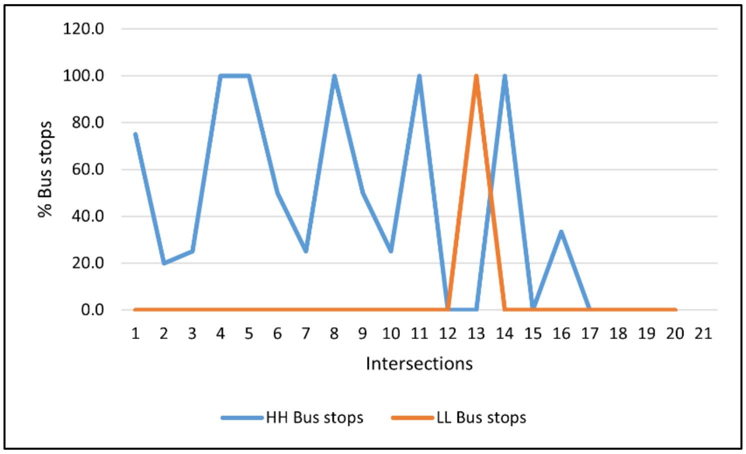

Figure 5.14: Line graph of bus stops

Parks in this study were defined as a green space for human recreation. In this study, parks were treated as an independent variable not inclusive of institutions because more trips are generated on the weekend, unlike governmental and healthcare institutions that cater to the public's needs for most of the weekday. Figure 5.15 shows that parks were mainly in the HH crash intersections than LL crash intersections.

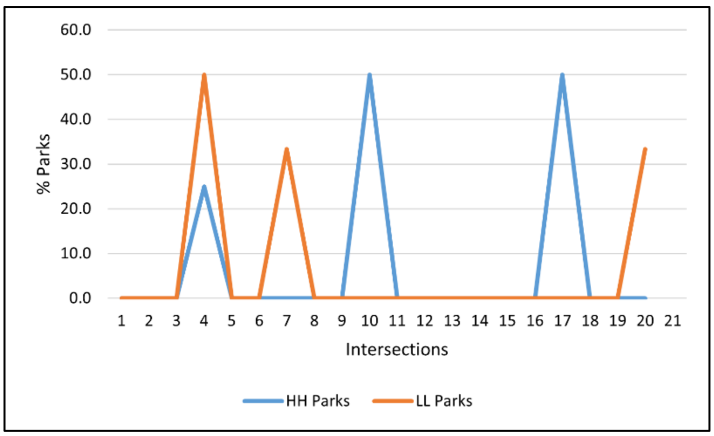

Figure 5.15: Line graph of parks

In this study, trees were defined as a plant with an elongated stem, supporting branches, and leaves. Surveying the built environment using GSV showed trees in single family residential only land use compared to commercial/institutional land use. Figure 5.16 shows that trees were mainly in the LL crash intersections than HH crash intersections.

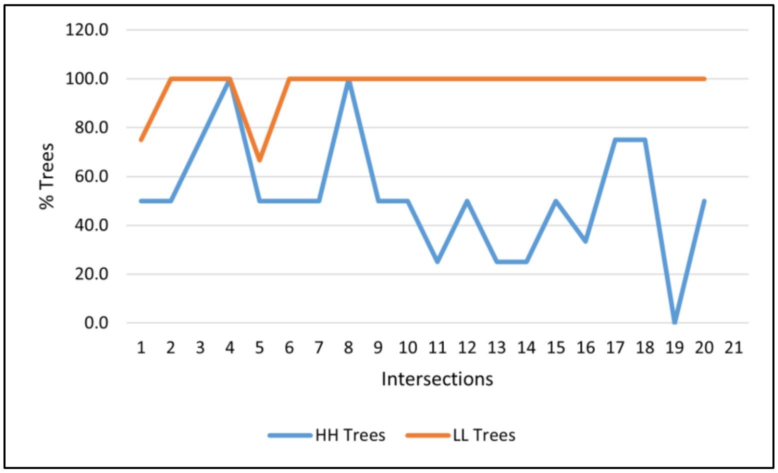

Figure 5.16: Line graph of trees

Sidewalks were defined as a dressed path for pedestrians. Figure 5.17 shows that the sidewalks were mainly in the HH crash intersections. The sidewalks were mostly on all intersection legs in HH crash areas, compared to the LL crash areas with sidewalks on some intersection legs. Some intersection approaches did not have sidewalks on all sides.

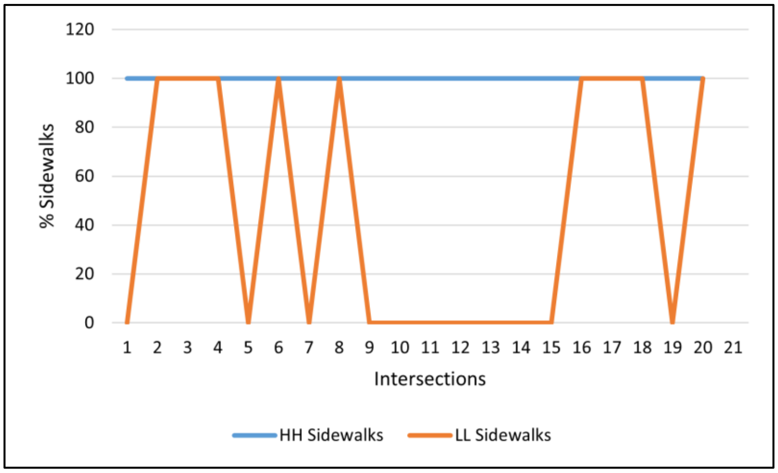

Figure 5.17: Line graph of sidewalks

Crosswalks in this study were defined as a designated path for pedestrians to cross the road. Surveying the built environment using GSV revealed crosswalks in commercial/institutional land use compared to single family residential only land use. Figure 5.18 shows that the crosswalks in the HH and LL crash intersections.

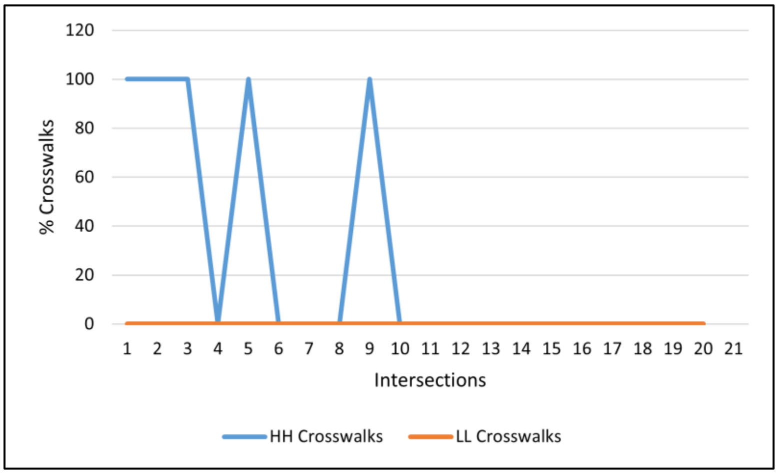

Figure 5.18: Line graph of crosswalks

Signalized intersections are intersections with signal controls or traffic lights. Surveying the built environment using GSV showed signalized intersections in commercial/institutional land use compared to single family residential only land use. Figure 5.19 shows that the signalized intersections in the HH and LL crash intersections.

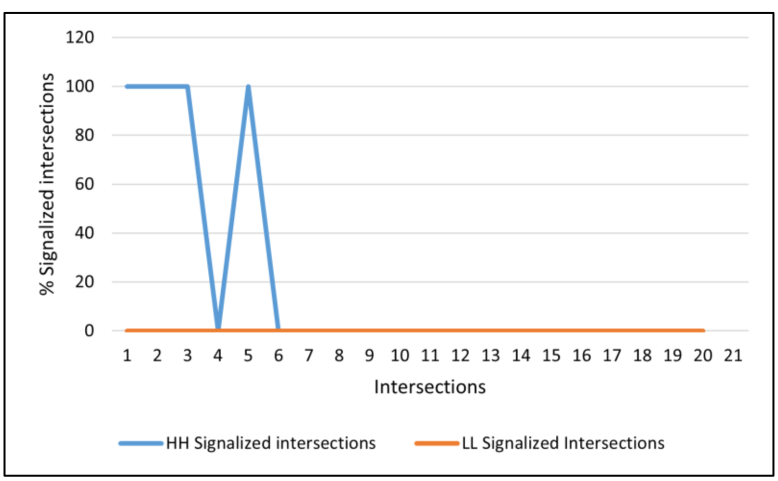

Figure 5.19: Line graph of signalized intersections

Priority intersections in this study refer to intersections where there are stop signs. Figure 5.19 shows a line diagram of priority intersections in the HH and LL crash intersections. Stops signs were mostly in single family residential land use compared to commercial/institutional land use.

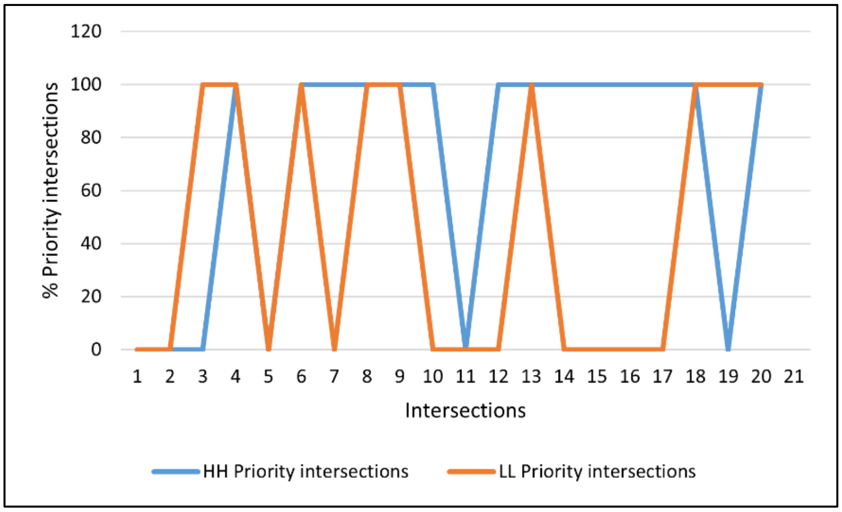

Figure 5.20: Line graph of priority intersections

5.4 Assessing the Relationship between Built Environment and Traffic Crashes

A dependent and independent variable were defined to assess the relationship between the built environment variables and motor vehicle traffic crashes. The intersection crash rate was the dependent variable, and the built environment variables and the poverty variable were the independent variables.

5.4.1 Dependent Variable – Intersection Crash Rate

The calculation of the intersection crash rate considers crash frequency (crashes per year) and vehicle exposure (traffic volumes). Intersection Crash rate is defined as crashes per 1 million entering vehicles (RMEV), such that

Intersection Crash Rate = C *1,000,000 /AADT *Y*365 where C = count of intersection traffic crashes within the duration analyzed, AADT= annual average traffic entering the intersection and Y= number of years analyzed (7 years for this study).

The MV intersection crash rate for the high and low cluster intersections calculated is shown in figure 5.21.

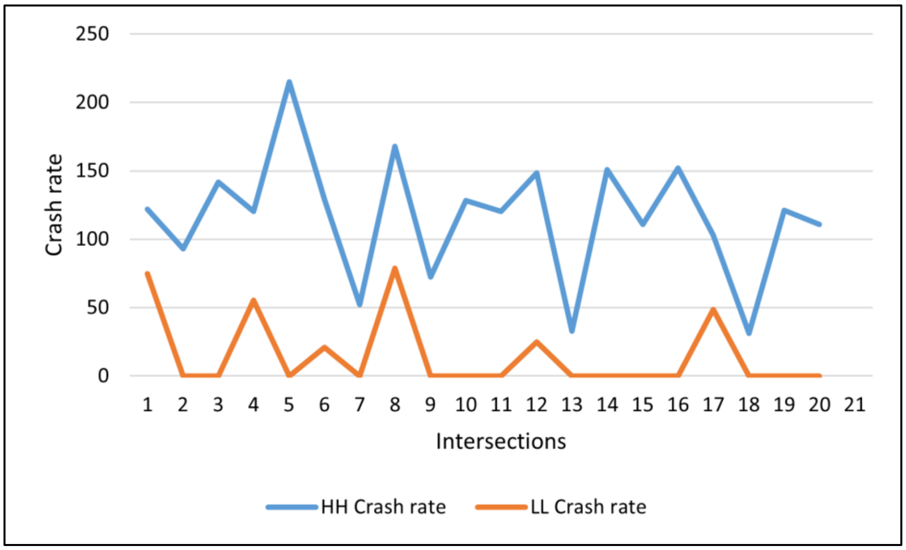

Figure 5.21: MV intersection crash rate

5.4.2 Independent Variables – Built Environment and Economic Variables

The independent variables are the built environment variables, which were obtained from previous studies, and a socioeconomic variable, specifically poverty. A total of 13 independent variables was used for the analysis, 12 built environment independent variables, and 1 socioeconomic variable. Google Street View was the tool used to obtain quantitative data on the built independent variables.

5.5 Regression Model

Three regression models were constructed, including the full multiple linear regression model, the stepwise regression model, and the Tobit model. Three regression models were necessary to understand better and examine which independent variables would be strong predictors of the occurrence of MV traffic crashes. The number of data points was 40, with 20 data points from intersections with high clusters of traffic crashes and another 20 data points from intersections with low clusters of traffic crashes. All three models were constructed in R Studio 4.0.1 statistical software.

5.5.1 Model 1- Full Multiple Linear Regression

Model 1 was constructed with 13 independent variables, including 12 built environment variables and 1 socioeconomic variable: poverty. Table 5.3 shows the results obtained. Poverty was the most significant variable in the model; hence there is a positive relationship between poverty and MV intersection traffic crashes. Signalized intersections and bus stops, however, were also quite significant in the model. The independent variables that were the least significant were parks and signage because these had the highest p-values.

The estimate represents the regression coefficient. While holding other independent variables in the model constant, this estimate shows the extent of change of the dependent variable's mean given a one-unit shift in each independent variable (Frost et al., 2020). A positive estimate indicates a positive relationship between the dependent and the independent variable, whilst a negative estimate represents a negative relationship. From the results in table 5.3, commercial/institutional land uses, signalized intersection, sidewalks, schools, signage, bus stops have a positive relationship with the occurrence of MV intersection traffic crashes. Single family residential, crosswalks, parks, on-street parking, and trees have a negative relationship with the MV intersection crash occurrence.

The standard error indicates the accuracy level of the predictions and shows the extent of the data points' distance from the regression line (Frost et al., 2020). The standard error has similar units as the dependent variable, the MV intersection crash rate. The standard error, together with the R-squared, are two ways to measure the regression model's goodness of fit. An increase in R-squared decreases the standard error and results in data points closer to the regression line. However, higher R-squared values do not provide information on the extent of distance the data points are from the regression line. Low values of the standard error indicate the distances between the data points, and the predicted values are smaller. The predicted value also refers to the fitted value. It can be defined as the model's prediction of the mean response value when you input the model's predictors' values. Table 5.3 shows that poverty, signalized intersections, bus stops, and commercial/institutional land use had the lowest standard errors implying they were closer to the regression line than the other independent variables.

R squared is the goodness-of-fit measure for linear regression models on a scale of 0 to 100%. It is a statistic that shows the percentage of the variance in the dependent variable that can be attributed to the independent variables (Frost et al., 2020). The R-squared is 74%, and the adjusted R-squared is 61%. For this model, the adjusted R-squared is more useful for analysis due to a higher number of independent variables. Therefore, about 61% of the total variation in the MV intersection crash rate is explained by the regression.

The t value shows the test statistic, with a bigger t value indicating the absence of a null hypothesis. Poverty, signalized intersections, and bus stops had the highest t values hence were against the null hypothesis. P-values are obtained based on the assumption that the null hypothesis is true, with lower p-values indicating a rejection. The results shown in table 5.3

indicate that poverty, bus stops, and signalized intersections had the lowest p-values hence were somewhat statically significant, and the null hypothesis was absent.

The validity of the model was also examined using the standard residual plot shown in figure 5.22, which has residuals spread around the horizontal line without a specific pattern hence an indicator of the absence of a non-linear relationship.

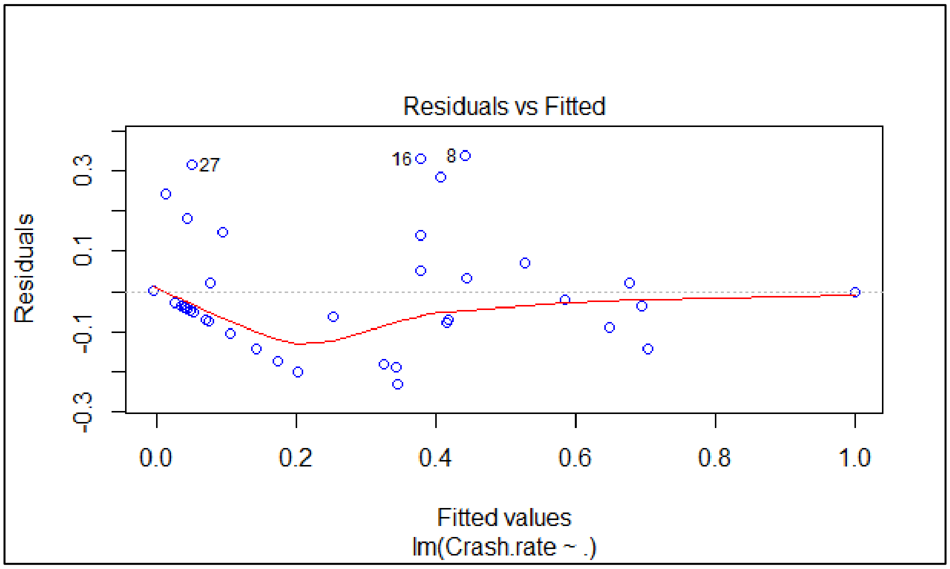

Figure 5.22: Residual plot

The results of the model are somewhat satisfactory; although 40 data points were used, these data points were from exploratory data analysis that resulted in the selection of statistically significant data points. Perhaps more data points would have strongly predicted the significance of commercial/institutional land uses, bus stops, and signalized intersections; therefore, a stepwise regression and Tobit model was explored for further understanding.

Table 5.3: Full multiple linear regression

Variables	Estimate	Standard error	t value	Pr (>| t |)
Intercept	0.080	0.168	0.464	0.647
Poverty	0.386	0.177	2.175	0.039*
Commercial/Institutional land uses	0.167	0.173	0.965	0.344
Single family residential	-0.050	0.112	-0.469	0.643
Signalized intersection	0.378	0.220	1.717	0.098 .
Sidewalks	0.035	0.089	0.400	0.692
Crosswalks	-0.464	0.277	-1.674	0.106
Priority Intersection	-0.030	0.079	-0.390	0.699
Schools	0.361	0.331	1.092	0.285
Parks	-0.012	0.270	-0.044	0.966
Signage	0.021	0.156	0.136	0.893
On-street Parking	-0.014	0.085	-0.164	0.871
Bus stops	0.193	0.099	1.932	0.064 .
Trees	-0.026	0.093	-0.277	0.784

Residual standard error: 0.186 on 26 degrees of freedom

Multiple R-squared: 0.7403, Adjusted R-squared: 0.6104

F-statistic: 5.701 on 13 and 26 DF, P-value: 8.35e-05

5.5.2 Model 2- Stepwise Multiple Linear Regression

Model 2 was constructed on the premise of a stepwise regression process. Table 5.4 shows the commercial/institutional land uses, signalized intersections, poverty, and bus stops are statistically significant independent variables with low p-values. The model indicates that crosswalks have a negative relationship with the MV intersection crash rate. The contribution of the independent variables to the MV intersection crash rate occurrence depends on the magnitude of the variable; that is, the bigger its magnitude, the higher its contribution.

Table 5.4: Stepwise multiple linear regression

Variables	Estimate	Standard error	t value	Pr (>| t |)
Intercept	-0.008	0.464	-0.163	0.872
Poverty	0.417	0.134	3.107	0.004**
Commercial/Institutional land uses	0.254	0.071	3.577	0.001 **
Signalized intersection	0.405	0.203	2.257	0.031*
Crosswalks	-0.454	0.202	-2.236	0.032*
Bus Stops	0.183	0.077	2.370	0.024 **

Residual standard error: 0.165 on 34 degrees of freedom

Multiple R-squared: 0.7195, Adjusted R-squared: 0.6782

F-statistic: 17.44 on 5 and 34 DF, P-value: 1.524e -08

5.5.3 Model 3- Tobit Model

Tobit model, a censored regression model, was employed because crash rates for low-value clusters of traffic crashes were mostly zero. The MV intersection crash rates were continuous and non-negative, hence the Tobit model's suitability for regression analysis. The model shows the smallest p-value for the intercept as compared to model 1 and model 2. A small p-value of the intercept represents a rejection of the null hypothesis. The model shows that poverty, commercial/institutional land uses, and bus stops are statistically significant. Commercial/institutional land uses and bus stops have a positive relationship with MV crashes at intersections whilst poverty is positively associated. Commercial/institutional land uses had a higher estimate than bus stops from the results, hence contributing more to the MV intersection crash rate.

Table 5.5: Tobit model

Variables	**Estimate**	**Standard error**	**t value**	**Pr (>| t |)**
Intercept	-0.167	0.076	-2.181	0.0292*
Poverty	0.507	0.179	2.830	0.0046**
Commercial/Institutional land uses	0.388	0.100	3.850	0.0001***
Signalized intersections	0.379	0.233	1.628	0.1035
Crosswalks	-0.437	0.264	-1.656	0.0976 .
Bus stops	0.197	0.101	1.946	0.051.
logSigma	-1.541	0.151	-10.192	<2e-16***
Log-likelihood: -5.8669 on 7 Df				

CHAPTER 6. CONCLUSIONS

6.1 Research Findings

The research study employs an exploratory spatial data analysis using Geoda, a virtual audit using GSV, and statistical analysis using the multiple linear regression and Tobit model to investigate the relationship between the built environment and MV intersection traffic crashes. Research results from the regression models show that crosswalks, single-family residential, and priority intersections have a negative relationship with the occurrence of MV traffic crashes at intersections. Single family residential only land use rarely serves as an attraction to different trip purposes hence its inclination to have low occurrences of MV intersection traffic crashes.

Surveying the built environment revealed that priority intersections were more likely found in single family residential than in commercial/institutional land uses. Additionally, the bus stops and signalized intersections were most likely characteristics of intersections with varying land use mix than in monotonous land use like single family residential only.

The study results from the stepwise multiple linear regression model and Tobit model in table 5.4 and table 5.5 indicate that commercial/institutional land uses, bus stops, and signalized intersections have a positive and significant relationship with the occurrence of MV traffic crashes at intersections. Therefore, answering the research question, what visual elements of the built environment characterize the high and low clusters of MV traffic crash intersections? An example of an intersection within the high value clusters of MV traffic crash intersections is 19th Street and Forest Avenue, which happens to be a signalized intersection with crosswalks on all intersection legs. Figure 6.1 highlights a senior living institution; figure 6.2 highlights an elementary institution, the Martin Luther King Jr. elementary school; figure 6.3 highlights a single-family residential unit, and figure 6.4 highlights a bus stop. 19th Street and Forest Avenue have three built environment variables: the existence of an institution, traffic signal, and bus stop determined as statistically significant from the regression modeling.

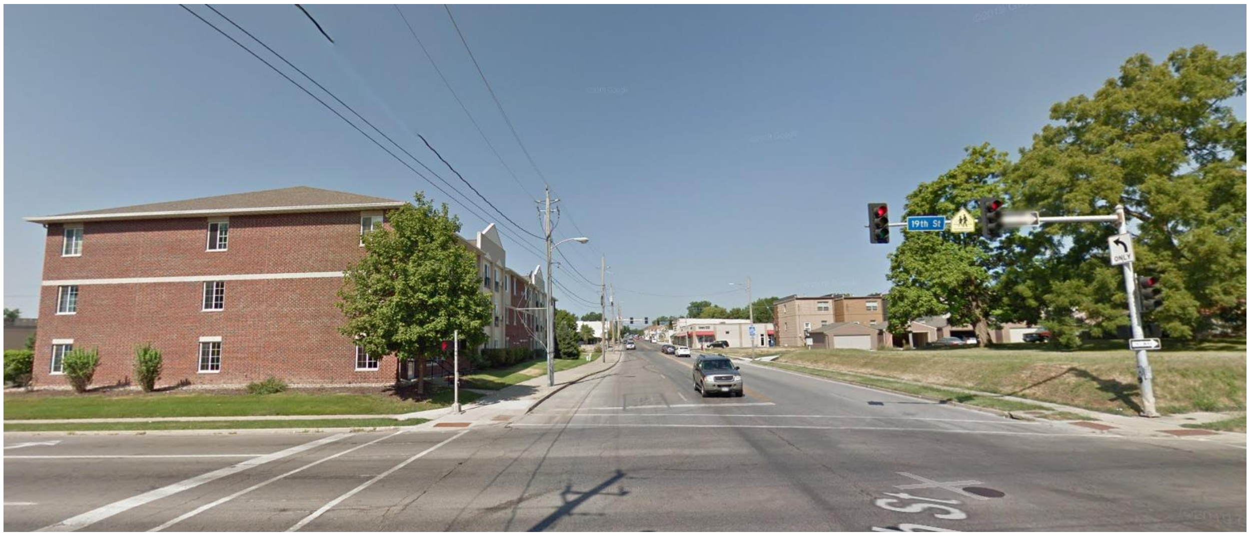

Figure 6.1: 19th Street & Forest Avenue (East direction)


Figure 6.2: 19th Street &Forest Avenue (West direction)

Figure 6.3: 19th Street and Forest Avenue (South direction)

Figure 6.4: 19th Street and Forest Avenue (North direction)

An example of an intersection within the low value clusters of MV traffic crash intersections is Forest Avenue & Forestdale Drive. Figure 6.5 highlights a single family residential, figure 6.6 shows a stop sign in the North direction, and figure 6.7 shows on-street parking and trees at the intersection. Forest Avenue & Forestdale Drive has two built environment variables: single family residential and stop sign (priority intersection) that had a negative relationship with MV intersection crash rate from the results of model 1 (full multiple linear regression model).

Figure 6.5: Forest Avenue & Forestdale Drive (West direction)

Figure 6.6: Forest Avenue & Forestdale Drive (North direction)

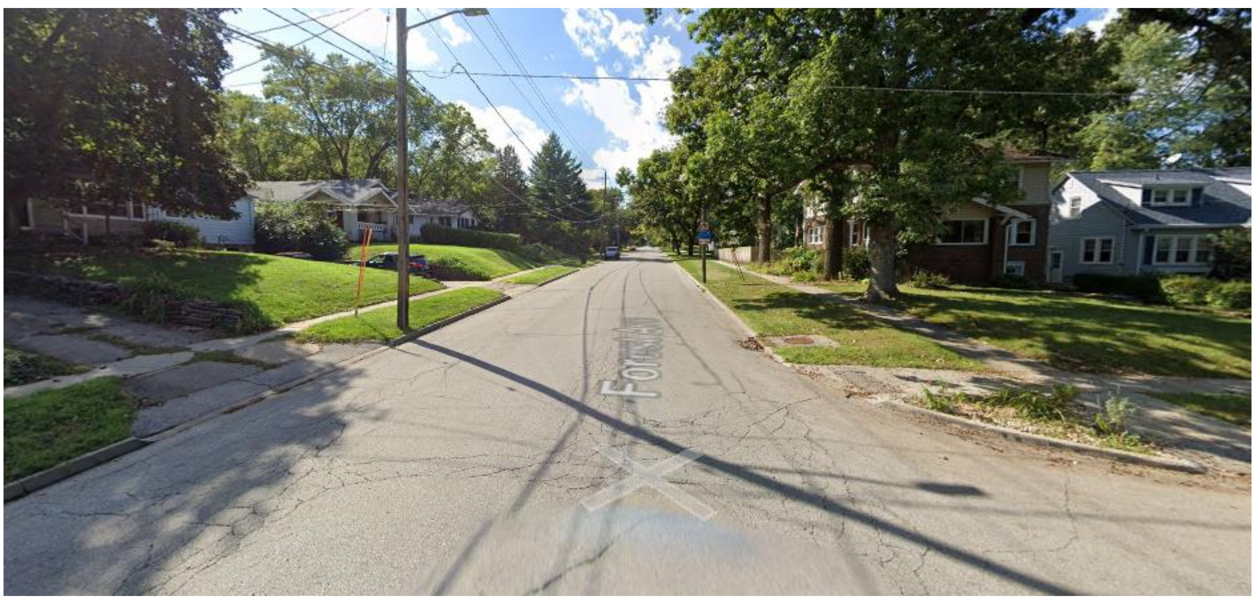

Figure 6.7: Forest Avenue & Forestdale Drive (East direction)

The study results in table 6.1 shows the intersections where there are high and low value clusters of urban traffic crashes during the period 2013-2019. This answers the research question, where are the locations of high and low clusters of urban MV traffic crash intersections during 2013-2019 in the City of Des Moines?

The study showed that the socioeconomic variable, poverty, has a positive relationship with the occurrence of MV intersection crashes. The relationship is not causal since alleviating poverty might not necessarily reduce MV intersection crashes for sure. Still, it sheds light on the socioeconomic status of high value cluster MV intersection crash neighborhoods. Therefore, answering the research question, what is the relationship between high and low clusters of MV traffic crash intersections and the level of poverty in the neighborhoods they are located? Figure 5.5 shows high value clusters of traffic crashes in low-income neighborhoods and low value clusters of traffic crashes in high-income neighborhoods.

Additionally, a high occurrence of MV traffic crashes occurred in the low-income areas of the following neighborhoods Merle Hay, Drake, Highland Park, Prospect Park, King Irving, Cheatom Park, Capitol Park, Martin Luther Jr. Park. The low occurrence of motor vehicle crashes occurred in the high-income areas of the following neighborhoods, Waveland Park, Lower Beaver, Beaverdale, Prospect Park, Chautauqua Park, Merle Hay, and Union Park.

Table 6.1: High and low clusters of traffic crash intersections

High Clusters of Traffic Crash Intersections	**Low Clusters of Traffic Crash Intersections**
Martin Luther King Jr Parkway & Clark Street	60th Street & Sheridan Avenue
19th Street & Carpenter Avenue	4th Street & Orchard Avenue
19th Street & Forest Avenue	47th Street & Beaver Crest Drive
6th Avenue & Indiana Avenue	55th Street & College Avenue
2nd Avenue & University Avenue	57th Street & College Avenue
3rd Street & Euclid Avenue	45th Street & Urbandale Avenue
York Street & E University Avenue	37th Street & Washington Avenue
31st Street & Brattleboro Avenue	Forest Avenue & Forestdale Drive
57th Street & Douglas Avenue	Beaver Avenue & Northwest Drive
E 14th Street & Fremont Street	E 18th Street & Mattern Avenue
6th Avenue & Ascension Street	Dixon Street & Mattern Avenue
3rd Street & University Avenue	Avalon Road & Burlington Terrace
1st Street & E Euclid Avenue	Payne Road & 26th Street
York Street & E Euclid Avenue	27th Street & Sheridan Avenue
E 14th Street & E Douglas Avenue	36th Street and Davisson Road
E 14th Street & E Oak Park Avenue	Sheridan Avenue & 42nd Street
E 15th Street & E University Avenue	Sheridan Avenue & 62nd Street
7th Street & University Avenue	New York Avenue & 61st Street
Merle Hay Road & Sutton Drive	60th Street & Franklin Avenue
Johnson Court & E 14th Street	California Drive & California Drive

6.2 Recommendations

The results of this study are beneficial for future land use planning and policies that will prioritize safety in transportation planning. Urban planners should examine commercial/institutional land use with a critical look to ensure safety conscious planning. Even though commercial/institutional land use existed in the hotspot's intersections, the existence of single-family residential contributed to a land use mix. Land use mix has been argued by urban planners to improve the quality of life.

Mixed land uses were strongly advocated by Jane Jacobs, who opposed modernist planning. Modernist planning is planning characterized by order, functionality, and zoning. "A modern city lives by the straight line, inevitably for the construction of buildings, sewers and tunnels, highways and pavements. The circulation of traffic demands the straight line; it is the proper thing for the heart of the city" (Corbusier, 1929). The zoning system is regarded as significant in modern planning for the city's future health, which is separating the city into zones or districts in which the height of buildings, the number of stories, and the building area are clearly specified (Marsh & Ford, 1909). Jacobs blames modern planning and zoning for contributing to dull undiversified city sidewalks, which leads to people feeling insecure and unsafe in the streets. High dwelling densities have a terrible reputation in orthodox planning and housing theory, but Jacobs opposes the theory that it leads to every kind of difficulty and failure. Jacobs argues that the supposed connection between high densities and danger, or high densities and slums in cities merely is inaccurate, as anyone who troubles to look at real cities can see (Jacobs, 1961).

Jacobs, however, did not pay much thought to high densities and traffic safety. High densities accompany public transit and high trip attractions. From this study, land use zoned to be a single family residential only had low value crash rates than land use with a mix of commercial/institutional and single family residential that was typically dense. It can be said that high density areas have failed with regards to traffic safety. Therefore, future land use and

zoning decisions should consider a pretty balance of high and low densities in a designated space so that trips are balanced and not concentrated in dense areas where there is a likelihood of MV intersection traffic crashes. More so, urban development and land use decisions should be made not only by urban planners but also by transportation planners, combining expertise from both professions, which will improve safety in the transportation planning process.

Furthermore, this study's results are useful in identifying the impact of the built environment in neighborhoods with a high traffic crashes occurrence, which can serve as a vital source of information on making decisions on transportation investments and long-range transportation plans. The regression model can be incorporated into a software tool for ease of use, where planners can enter in built environment variables as parameters. The software tool will predict MV intersection traffic crash rates for decision-making purposes, thereby promoting better incorporation of the community character into the transportation planning process.

Additionally, traffic calming measures can be implemented at intersection approaches with a high occurrence of traffic crashes. Traffic calming measures are the implementation of speed reduction mechanisms by altering the roadway configuration. Examples of traffic calming measures include chokers, chicanes, speed humps, lane shifts, roundabouts, etc. In conclusion, a combination of traffic calming measures should be prioritized in high crash neighborhoods in the City of Des Moines.

The location of bus stops and the optimal distance of bus stops from intersection approaches should be evaluated. The findings can be adopted into a policy framework for bus stop locations that consider the suitable distance and the land use mix at the intersection approach. Lastly, traffic signal configuration considering a land use mix environment should be examined by researchers; that is, traffic engineers need to take a critical look at the operational performance of signal configurations based on the land use mix.

6.3 Limitations and Future Research

In the future, the sociodemographic factors (age group, race, and ethnicity) of drivers where statistically significant urban intersection traffic crashes occurred can be studied. The extent of damage of crashes could be analyzed in the future; this includes (fatalities, serious injury, minor injury, and property damage only). This will enable the identification of intersections where fatalities occur most, hence concentrating efforts in such intersections. Additionally, the study can be expanded to include bicycle and pedestrian crashes; that is, the relationship between the built environment and bicycle and pedestrian crashes can be explored.

This research did not consider the characteristics of roads, such as speed limits. The speed limits of high and low clusters of urban traffic crash intersections can be considered in regression modeling.

Also, concerning this project's scalability, one limitation identified is the manual analysis of GSV images to identify built environment variables such as bus stops, signalized, and priority intersections. In the future, it will be necessary to overcome this bottleneck through automation by employing tools such as GSV API for automatic retrieval of images by coordinate points and using object recognition models trained on the built environment variables to label these images automatically.

In the future, a data visualization tool can be developed for managing traffic safety, which can be used at the county level, city level, and regional level. The tool can be developed to generate a report on intersection crashes resulting from land use decisions in an existing development or a new undeveloped area. This data visualization would help policymakers and transportation safety professionals in improving safety at intersections. This tool will ultimately help urban planners in land use planning and zoning decisions and be useful to transportation safety professionals in reducing MV traffic crashes at intersections.

REFERENCES

Anastasopoulos, P. C., Tarko, A. P., & Mannering, F. L. (2008). Tobit analysis of vehicle accident rates on interstate highways. Accident Analysis & Prevention, 40(2), 768-775. doi:10.1016/j.aap.2007.09.006.

Anselin, L. (2020). Contiguity-based spatial weights. Retrieved from https://geodacenter.github.io/workbook/4a_contig_weights/lab4a.html

Anselin, L. (2020). Distance-band spatial weights. Retrieved from https://geodacenter.github.io/workbook/4b_dist_weights/lab4b.html

Anselin, L. (2020). Global spatial autocorrelation. Retrieved from https://geodacenter.github.io/workbook/5a_global_auto/lab5a.html

Anselin, L.(1995). Local Indicators of spatial association-LISA. Geographical Analysis 27(2):92-116.

Badland, H. M., Opit, S., Witten, K., Kearns, R. A., & Mavoa, S. (2010). Can Virtual Streetscape Audits Reliably Replace Physical Streetscape Audits? Journal of Urban Health, 87(6), 1007-1016. doi:10.1007/s11524-010-9505-x.

Blincoe, L. J., Miller, T. R., Zaloshnja, E., & Lawrence, B. A. (2015, May). The economic and societal impact of motor vehicle crashes, 2010. (Revised) (Report No. DOT HS 812 013). Washington, DC: National Highway Traffic Safety Administration.

Chin, H. C., & Quddus, M. A. (2003). Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. Accident Analysis & Prevention, 35(2), 253-259.

Clarke, P., Ailshire, J., Melendez, R., Bader, M., & Morenoff, J. (2010). Using Google Earth to conduct a neighborhood audit: Reliability of a virtual audit instrument. Health & Place, 16, 1224-1229.

Clifton, K. J., Burnier, C. V., & Akar, G. (2009). Severity of injury resulting from pedestrian–vehicle crashes: What can we learn from examining the built environment? Transportation Research Part D: Transport and Environment, 14(6), 425–436. doi: 10.1016/j.trd.2009.01.001.

Corbusier, L. (1929). The city of tomorrow and its planning. Trans. 8th French Edition. Etchells, Frederick. Urbanisme. London: The Productional Press.

Cottrill, C. D., & Thakuriah, P. (. (2010). Evaluating pedestrian crashes in areas with high low-income or minority populations. Accident Analysis &; Prevention, 42(6), 1718-1728. doi:10.1016/j.aap.2010.04.012.

Dezman, Z., Andrade, L. D., Vissoci, J. R., El-Gabri, D., Johnson, A., Hirshon, J. M., & Staton, C. A. (2016). Hotspots and causes of motor vehicle crashes in Baltimore, Maryland: A geospatial analysis of five years of police crash and census data. Injury, 47(11), 2450–2458. doi: 10.1016/j.injury.2016.09.002.

Dumbaugh, E., & Li, W. (2010). Designing for the Safety of Pedestrians, Cyclists, and Motorists in Urban Environments. Journal of the American Planning Association, 77(1), 69–88. doi: 10.1080/01944363.2011.536101.

Esri (2020) How Spatial Autocorrelation (Global Moran's I) works. Retrieved from https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm

Elvik, R. (2008). The predictive validity of empirical Bayes estimates of road safety. Accident Analysis & Prevention, 40(6), 1964-1969. doi:10.1016/j.aap.2008.07.007.

Esri. (2020). Spatial Join (Analysis). Retrieved from https://pro.arcgis.com/en/pro-app/tool-reference/analysis/spatial-join.htm

Ewing, R. (2015). Pedestrian Safety and the Built Environment. Journal of Planning Literature, 30(4), 377-392. doi:10.1177/0885412215595438.

Ewing, R., & Dumbaugh, E. (2009). The Built Environment and Traffic Safety. Journal of Planning Literature, 23(4), 347–367. doi: 10.1177/0885412209335553.

Federal Safety Net. (2020). Poverty Definition. Retrieved from http://federalsafetynet.com /poverty-definition.html

Frost, J., Laurie, Mondal, B., Katja, Lee, J., Renu & Dubey, D. (2020). How To Interpret R-squared in Regression Analysis. Retrieved from https://statisticsbyjim.com/regression/interpret-r-squared-regression/

Gladhill, K., & Monsere, C. M. (2012). Exploring traffic safety and urban form in Portland, Oregon. Transportation research record, 2318(1), 63-74.

Handy, S. L., Boarnet, M. G., Ewing, R., & Killingsworth, R. E. (2002). How the built environment affects physical activity. American Journal of Preventive Medicine, 23(2), 64-73. doi:10.1016/s0749-3797(02)00475-0.

Henningsen, A. (2010). Estimating censored regression models in R using the censReg Package. R package vignettes collection, 5(2), 12.

Hoyert, D. L., Kung, H.-C., & Smith, B. L. (2005). Deaths: preliminary data for 2003. National vital statistics reports, 53(15), 1-48.

Huang, Y., Wang, X., & Patton, D. (2018). Examining spatial relationships between crashes and the built environment: A geographically weighted regression approach. Journal of Transport Geography, 69, 221-233. doi:10.1016/j.jtrangeo.2018.04.027.

Humphrey N.P. (2005) Does the Built Environment Influence Physical Activity? Examining the Evidence. Washington DC, Transportation Research Board and Institute of Medicine.

Iowa Department of Transportation (2020). Iowa Crash Analysis Tool (ICAT). Retrieved from https://icat.iowadot.gov/

Iowa Department of Transportation (2020). Iowa Traffic Data. Retrieved from https://iowadot.maps.arcgis.com/apps/MapSeries/index.html?appid=0cce99afb78e4d3b9b24f8263717f910

Jacobs, J. (1961). The Death and Life of Great American Cities. New York: Random House.

K. L. Wolf and N. J. Bratton, "Urban trees and traffic safety: Considering U.S. roadside policy and crash data," Arboriculture Urban Forestry, vol. 32, no. 4, pp. 170–179, Jul. 2006.

Kaygisiz, Ö, Senbil, M., & Yildiz, A. (2017). Influence of urban built environment on traffic accidents: The case of Eskisehir (Turkey). Case Studies on Transport Policy, 5(2), 306-313. doi:10.1016/j.cstp.2017.02.002.

Kelly, C. M., Wilson, J. S., Baker, E. A., Miller, D. K., & Schootman, M. (2013). Using Google Street View to audit the built environment: inter-rater reliability results. Annals of Behavioral Medicine, 45(suppl_1), S108-S112.

Kim, K., I. M. Brunner, and E. Y. Yamashita. Influence of Land Use, Population, Employment, and Economic Activity on Accidents. In Transportation Research Record: Journal of the Transportation Research Board, No. 1953, Transportation Research Board of the National Academies, Washington, D.C., 2006.

Marsh, Benjamin Clarke, and George B. Ford. An Introduction to City Planning: Democracy's Challenge to the American City. Memphis, TN: General Books LLC, 1909.

Meyer, M., & Miller, E. (2001). Urban Transportation Planning: A Decision-Oriented Approach.

Miranda-Moreno, L. F., Morency, P., & El-Geneidy, A. M. (2011). The link between built environment, pedestrian activity, and pedestrian–vehicle collision occurrence at signalized intersections. Accident Analysis & Prevention, 43(5), 1624-1634. doi:10.1016/j.aap.2011.02.005

Mooney, S. J., Bader, M. D. M., Lovasi, G. S., Neckerman, K. M., Teitler, J. O., & Rundle, A. G. (2014). Validity of an Ecometric Neighborhood Physical Disorder Measure Constructed by Virtual Street Audit. American Journal of Epidemiology, 180(6), 626-635.

Morency, P., Gauvin, L., Plante, C., Fournier, M., & Morency, C. (2012). Neighborhood Social Inequalities in Road Traffic Injuries: The Influence of Traffic Volume and Road Design. American Journal of Public Health, 102(6), 1112-1119. doi:10.2105/ajph.2011.300528.

Odgers, C. L., Caspi, A., Bates, C. J., Sampson, R. J., & Moffitt, T. E. (2012). Systematic social observation of children's neighborhoods using Google Street View: a reliable and cost-effective method. Journal of Child Psychology and Psychiatry, 53(10), 1009-1017.

Ouyang, Y., & Bejleri, I. (2014). Geographic Information System–Based Community-Level Method to Evaluate the Influence of Built Environment on Traffic Crashes. Transportation Research Record: Journal of the Transportation Research Board, 2432(1), 124–132. doi: 10.3141/2432-15.

Park, M., & Lee, D. (2017). Analysis of Severe Injury Accident Rates on Interstate Highways Using a Random Parameter Tobit Model. Mathematical Problems in Engineering, 2017, 1-6. doi:10.1155/2017/7273630.

Pulugurtha, S. S., Duddu, V. R., & Kotagiri, Y. (2013). Traffic analysis zone level crash estimation models based on land use characteristics. Accident Analysis & Prevention, 50, 678-687. doi:10.1016/j.aap.2012.06.016.

Rundle, A. G., Bader, M. D., Richards, C. A., Neckerman, K. M., & Teitler, J. O. (2011). Using Google Street View to audit neighborhood environments. American journal of preventive medicine, 40(1), 94-100.

Soltani, A., & Askari, S. (2017). Exploring spatial autocorrelation of traffic crashes based on severity. Injury, 48(3), 637-647.

Statistics solutions (2020). Assumptions of multiple linear regression. Retrieved from https://www.statisticssolutions.com/assumptions-of-multiple-linear-regression/

Stoker, P., Garfinkel-Castro, A., Khayesi, M., Odero, W., Mwangi, M. N., Peden, M., & Ewing, R. (2015). Pedestrian safety and the built environment: a review of the risk factors. Journal of Planning Literature, 30(4), 377-392.

Sze, N., & Wong, S. (2007). Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. Accident Analysis & Prevention, 39(6), 1267-1278. doi:10.1016/j.aap.2007.03.017.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. Econometrica 26 (1), 24–36.

Li, Z. (2018). Transportation asset management: Methodology and applications. Boca Raton, FL: CRC Press, Taylor & Francis Group.

National Safety Council. (2020). Retrieved October 04, 2020, from https://injuryfacts.nsc.org/motor-vehicle/overview/introduction/

U.S. Census Bureau. (2020). QuickFacts: Des Moines city Iowa. Retrieved from https://www.census.gov/quickfacts/desmoinescityiowa

Wang, Y., & Kockelman, K. M. (2013). A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. Accident Analysis & Prevention, 60, 71-84. doi:10.1016/j.aap.2013.07.030.

Xie, Z., & Yan, J. (2008). Kernel density estimation of traffic accidents in a network space. Computers, environment and urban systems, 32(5), 396-406.

Zahabi, S. A., Strauss, J., Manaugh, K., & Miranda-Moreno, L. F. (2011). Estimating Potential Effect of Speed Limits, Built Environment, and Other Factors on Severity of Pedestrian and Cyclist Injuries in Crashes. Transportation Research Record: Journal of the Transportation Research Board, 2247(1), 81-90. doi:10.3141/2247-10.

APPENDIX A. GSV QUESTIONNAIRE

The questionnaire will examine the visual elements in the built environment that may affect the frequency of motor vehicle traffic crashes based on extensive literature review.

What are the street names at the intersection?

____

What is the date of imagery?

____

What is the observed latitude on google street view?

____

What is the observed longitude on google street view?

____

What is the direction of the street intersection?

____

Are single family residential buildings visible from the intersection?

☐ Yes ☐ No

Are schools visible from the intersection?

☐ Yes ☐ No

Are industrial areas visible from the intersection?

☐ Yes ☐ No

Are institutional areas visible from the intersection?

☐ Hospitals

☐ Orphanages

☐ Clinics

- [ ] Church
- [ ] Senior living family
- [ ] Government institution
- [ ] Financial
- [ ] NGO’s/ Community Center

Any additional or other institutional areas not listed

____

Are commercial areas visible form the intersection?

- [ ] Office complexes
- [ ] Shopping mall
- [ ] Service stations
- [ ] Restaurants
- [ ] Retail spaces
- [ ] Visitor lodging
- [ ] Grocery Store
- [ ] Fast Food/ Ice Cream
- [ ] Laundry
- [ ] Car wash
- [ ] Home Business
- [ ] Pharmacy
- [ ] Beauty Service
- [ ] Warehouse/Storage

Any additional or other commercial areas not listed?

____

Is there a research district area visible from the intersection that is not listed?

☐ Yes ☐ No

Are parks visible form the intersection?

☐ Yes ☐ No

Are signages visible from the intersection?

☐ Yes ☐ No

Is there a crosswalk at the intersection?

☐ Yes ☐ No

How many legs does the intersection have?

☐ 2

☐ 3

☐ 4

☐ 5

Are sidewalks visible at the intersection?

☐ Yes ☐ No

Is on-street parking visible at the intersection?

☐ Yes ☐ No

Are bus stops visible from the intersection?

☐ Yes ☐ No

Are trees visible from the intersection?

☐ Yes ☐ No

Are traffic signals visible from the intersection?

☐ Yes ☐ No

Are stop signs visible from the intersection?

☐ Yes ☐ No

APPENDIX B. PYTHON CODE

```
# program to merge intersections of close proximity

def distance(x, y):
    # Return the Euclidean distance between points x and y
    return ((x[0] - y[0])**2 + (x[1] - y[1])**2) ** 0.5

# threshold values
threshold = 20
# open original and new file
new = open("Filtered_intersection_points.csv", "a", encoding = "utf-8")
original = open("All_intersection_points.csv", "r", encoding = "utf-8")

# read in all points
lines = original.readlines()

# list for all visited points
visited = []


for line in lines:
    eachline = line.split(',')
    x_coord = float(eachline[6])
    y_coord = float(eachline[7])
    point = (x_coord,y_coord)
    count = 0
    for visit in visited:
        if distance(point,visit) <= threshold:
            count += 1
    if count == 0:
        new.write(line)
        visited += [point]
# close files
original.close()
new.close()
```